PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-271521

(43)Date of publication of application : 20.10.1995

(51)Int.CI. G06F 3/06
G06F 13/10

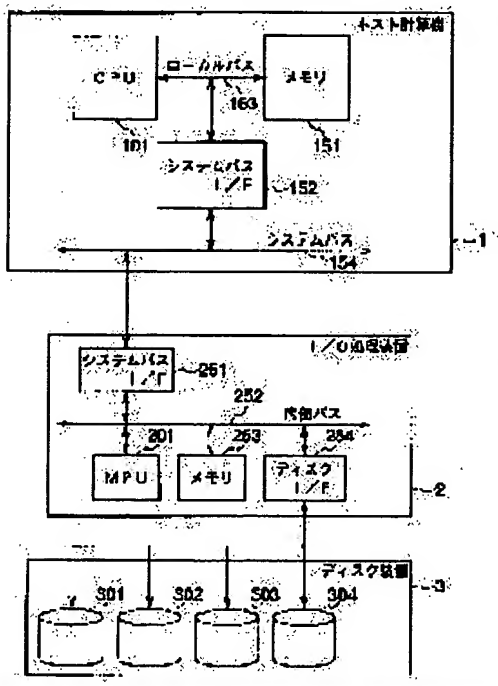
(21)Application number : 06-057197

(22)Date of filing : 28.03.1994

(71)Applicant : HITACHI LTD

(72)Inventor : MATSUNAMI NAOTO
SUGA MASAYUKI
KANEDA TAISUKE
YAGISAWA IKUYA
OEDA TAKASHI
ARAKAWA TAKASHI

(54) COMPUTER SYSTEM



(57)Abstract:

PURPOSE: To reduce the interruption frequency to a CPU to report the end of the disk access.

CONSTITUTION: A CPU 101 of a host 1 issues plural disk commands to an I/O processor 2 by a disk command management means independently of and in relation to each other. An MPU 201 of the processor 2 registers the relation among received disk access commands and issues it to a disk device 3 by a disk command management part 203. Then the MPU 201 issues an interruption to the CPU 101 and informs a fact that all related commands are finished after this fact is decided by the end report received from the device 3.

LEGAL STATUS

[Date of request for examination] 01.07.1998

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3209634

[Date of registration] 13.07.2001

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（Ａ）**　(11)特許出願公開番号

特開平7－271521

(43)公開日　平成７年(1995)10月20日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ０６Ｆ　3/06	５４０			
13/10	３４０　Ｂ	7368－5Ｂ		

審査請求　未請求　請求項の数13　ＯＬ　（全 23 頁）

(21)出願番号　特願平6－57197

(22)出願日　平成６年(1994)３月28日

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地

(72)発明者　松並　直人
神奈川県川崎市麻生区王禅寺1099　株式会社日立製作所システム開発研究所内

(72)発明者　菅　政之
神奈川県海老名市下今泉810番地　株式会社日立製作所オフィスシステム事業部内

(72)発明者　兼田　泰典
神奈川県川崎市麻生区王禅寺1099　株式会社日立製作所システム開発研究所内

(74)代理人　弁理士　富田　和子

最終頁に続く

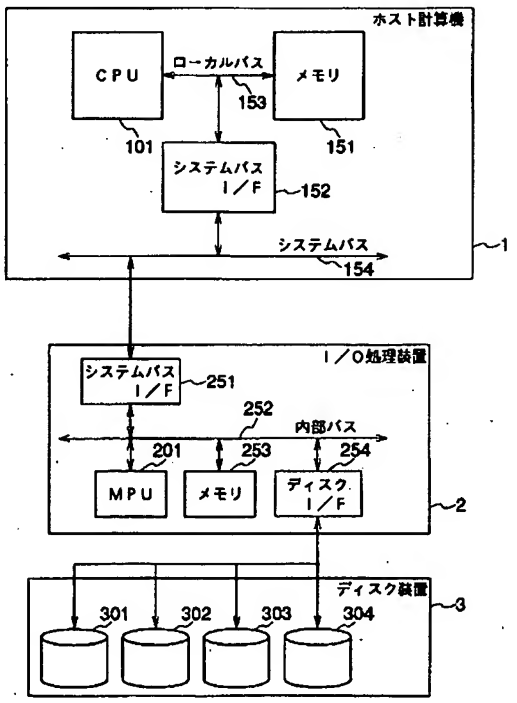
(54)【発明の名称】　計算機システム

(57)【要約】　（修正有）

【目的】ディスクアクセス終了を報告するＣＰＵへの割込み回数を削減する。

【構成】ホスト１のＣＰＵ１０１は、ディスクコマンド管理手段で複数のディスクコマンドを関連付けて独立にＩ／Ｏ処理装置２に発行する。Ｉ／Ｏ処理装置２のＭＰＵ２０１は、ディスクコマンド管理部２０３で、受け取ったディスクアクセスコマンド間の関連を登録しディスク装置３に発行し、ディスク装置３からの終了報告より、関連付けられたすべてのコマンドが終了したことを判定すると、ＣＰＵ１０１に割込みを発行し、関連付けられたコマンドすべてが終了したことを通知する。

図1



【特許請求の範囲】

【請求項１】割込み処理機構を備えたＣＰＵとメモリとを備えたホスト装置と、データの記録再生を行なう補助記憶装置と、ホスト装置と補助記憶装置との間の入出力を担う入出力処理装置とを備えた計算機であって、

前記ホスト装置のＣＰＵは、前記補助記憶装置のアクセスを指示する、複数のコマンドを、所定の条件に従いグループ化し、前記メモリに、各グループと当該グループに属するコマンドの対応を表す情報を記憶し、各コマンドに当該コマンドが属するグループを識別可能な識別情報を付加し、前記識別情報を付加した複数のコマンドを独立に前記入出力処理装置に渡す関連付け手段と、前記入出力処理装置より特定のグループについての処理が終了したことを割込みによって報告された場合に、前記メモリに記憶した各グループと当該グループに属するコマンドの対応を参照し、前記特定のグループに属する全てのコマンドに関する、所定の終了処理を行なう終了処理手段とを有し、

前記入出力処理装置は、渡された複数のコマンドに付加されている前記識別情報より、渡された各コマンドと当該コマンドが属するグループの対応を表す情報を作成して記憶する手段と、前記ＣＰＵより独立に発行された各コマンドの指示する補助記憶装置のアクセスを実行する手段と、記憶した各コマンドと当該コマンドが属するグループの対応に基づいて、１つのグループに属する全てのコマンドの指示するアクセスが終了した場合に、当該グループについての処理が終了したことを前記ＣＰＵに割り込みを用いて報告する手段とを有することを特徴とする計算機システム。

【請求項２】割込み処理機構を備えたＣＰＵとメモリとを備えたホスト装置と、補助記録媒体にデータの記録再生を行なう補助記憶装置と、補助記憶装置とホスト装置との間の入出力を担う入出力処理装置とを備えた計算機であって、

前記ＣＰＵは、前記補助記憶装置のアクセスを指示する、複数のコマンドを、所定の条件に従いグループ化し、前記メモリに、各グループと当該グループに属するコマンドの対応を表す情報を記憶し、前記複数のコマンドを独立に前記入出力処理装置に渡す関連付け手段と、前記入出力処理装置より特定のグループについての処理が終了したことを報告された場合に、前記メモリに記憶した各グループと当該グループに属するコマンドの対応を参照し、前記特定のグループに属する全てのコマンドに関する、所定の終了処理を行なう終了処理手段とを有し、

前記入出力処理装置は、前記ＣＰＵより独立に発行された各コマンドの指示する補助記憶装置のアクセスを実行する手段と、前記メモリに記憶された各グループと当該グループに属するコマンドの対応に基づいて、１つのグループに属する全てのコマンドの指示するアクセスが終了した場合に、当該グループについての処理が終了したことを前記ＣＰＵに割込みを用いて報告する手段とを有することを特徴とする計算機システム。

【請求項３】請求項１または２記載の計算機システムであって、

前記補助記憶装置は、複数のディスク装置から構成されるディスクアレイ装置であって、

前記ホスト装置のＣＰＵは、前記ディスクアレイ全体に対するアクセスを指示するディスクアレイコマンドより、ディスクアレイ装置を構成する個々のディスク装置へのコマンドを生成する手段と、ディスクアレイコマンドと個々のディスク装置へのコマンドの対応を前記メモリに記憶する手段とを有し、

前記関連付け手段は、一つの前記ディスクアレイコマンドに対応する個々のディスク装置へのコマンドを、一つの前記グループとし、

前記終了処理手段は、前記特定のグループに属する全てのコマンドに関する、所定の終了処理として、前記特定のグループに属するコマンドに対応して前記メモリに記憶されているディスクアレイコマンドの終了処理を行なうことを特徴とする計算機システム。

【請求項４】請求項１または２記載の計算機システムであって、

前記補助記憶装置は、データを格納する複数のデータ用ディスク装置と、前記複数のデータ用ディスク装置の同ディスク内アドレスの領域に記録されているデータ全体のパリティを記憶する１以上のパリティ用ディスク装置から構成されるＲＡＩＤ型のディスクアレイ装置であって、

前記ホスト装置のＣＰＵは、前記ディスクアレイ全体に対する新データの書き込みアクセスを指示するディスクアレイコマンドより、新データの前記データ用ディスク装置への書き込みアクセスを指示する新データ書き込み用コマンドと、前記新データの書き込みアクセスにより更新される、前記データ用ディスク装置内の現データの読み出しアクセスを指示する現データ読み出し用コマンドと、前記新データの書き込みアクセスにより更新される、前記データ用ディスク装置内のデータに対応する、前記パリティ用ディスク装置内の現パリティの読み出しアクセスを指示する現パリティ読み出し用コマンドと、前記新データの前記データ用ディスク装置への書き込みアクセスにより更新される、前記新データに対応する新パリティの前記パリティ用ディスク装置への書き込みアクセスを指示する新パリティ書き込み用コマンドを生成する手段と、

新パリティ生成するパリティ生成部とを有し、

前記関連づけ手段は、現データ読み出し用コマンドと前記現パリティ読み出し用コマンドとを一つの前記グループとし、現データ読み出し用コマンドと前記現パリティ読み出し用コマンドを前記入出力処理装置に渡し、その

後、前記パリティ生成手段によって、前記新パリティ書き込み用コマンドによって書き込む新パリティが生成されたら、新データ書き込み用コマンドと前記新パリティ書き込み用コマンドとを一つの前記グループとし、新データ書き込み用コマンドと前記新パリティ書き込み用コマンドとを前記入出力処理装置に渡し、
前記パリティ生成部は、前記入出力処理装置が、渡された現データ読み出し用コマンドに応じて前記ディスクアレイより読み出した前記現データと現パリティ用読み出し用コマンドに応じて前記ディスクアレイより読み出した現パリティと、新データ書き込み用コマンドによって書き込む新データとの排他的論理和を、前記新パリティ書き込み用コマンドによって書き込む新パリティとして生成することを特徴とする計算機システム。
【請求項５】請求項４記載の計算機システムであって、現データ読み出し用コマンドと前記現パリティ読み出し用コマンドとの前記グループと、新データ書き込み用コマンドと前記新パリティ書き込み用コマンドとの前記グループと、これらの二つのグループに対応するディスクアレイコマンドとの対応を示す情報を前記メモリに記憶する手段を有し、
前記終了処理手段は、前記特定のグループに属する全てのコマンドに関する、所定の終了処理として、前記メモリに前記特定のグループに対応して記憶されているディスクアレイコマンドに対応して記憶されている２つのグループについての処理が終了したことを、前記入出力処理装置より既に報告されているか否かを判定し、既に成されている場合に、当該ディスクアレイコマンドの終了処理を行なうことを特徴とする計算機システム。
【請求項６】請求項１または２記載の計算機システムであって、
前記ホスト装置のＣＰＵは、前記補助記憶装置に書き込むデータを一旦、メモリ上に設けたキャッシュ領域に書き込む手段と、所定の契機で、前記キャッシュ領域に書き込まれた複数のデータの前記補助記憶装置への書き込みアクセスを指示する複数のコマンドを生成する書き戻し要求部とを、さらに有し、
前記関連付け手段は、前記書き戻し要求部が、前記所定の契機で生成した複数のコマンドの一部もしくは全部を、一つの前記グループとすることを特徴とする計算機システム。
【請求項７】請求項６記載の計算機システムであって、前記書き戻し要求部は、所定の条件が満たされた場合に、前記生成した、前記キャッシュ領域に書き込まれた複数のデータの前記補助記憶装置への書き込みアクセスを指示する複数のコマンドに、グループ化を禁止する情報を付加する手段を有し、
前記関連付け手段は、前記書き戻し要求部が、前記所定の契機で生成した複数のコマンドのうち、グループ化を禁止する情報が付されているコマンドについてはグループ化の対象としないことを特徴とする計算機システム。
【請求項８】請求項１または２記載の計算機システムであって、
前記補助記憶装置は、データを格納する複数のデータ用ディスク装置と、前記複数のデータ用ディスク装置の同ディスク内アドレスの領域に記録されているデータ全体のパリティを記憶する１以上のパリティ用ディスク装置から構成されるＲＡＩＤ型ディスクアレイ装置であって、
前記ホスト装置のＣＰＵは、前記ディスクアレイ全体に対する新データの書き込みアクセスを指示するディスクアレイコマンドより、前記データ用ディスク装置への前記新データの書き込みアクセスを指示する新データ書き込み用コマンドと、前記新データの書き込みアクセスにより更新される、前記データ用ディスク装置内の現データの読み出しアクセスを指示する現データ読み出し用コマンドと、前記新データの書き込みアクセスにより更新される、前記データ用ディスク装置内のデータに対応する、前記パリティ用ディスク装置内の現パリティの読み出しアクセスを指示する現パリティ読み出し用コマンドと、前記新データの前記データ用ディスク装置への書き込みアクセスにより更新される、前記新データに対応する新パリティの前記パリティ用ディスク装置への書き込みアクセスを指示する新パリティ書き込み用コマンドを生成する手段と、
新パリティを生成するパリティ生成部とを有し、
前記関連付け手段は、現データ読み出し用コマンドを前記入出力処理装置に渡し、前記入出力処理装置によって現データが前記ディスクアレイより読み出されたら、現データ書き込み用コマンドを前記入出力処理装置に渡し、その後所定期間経過後に、複数の現パリティ読み出し用コマンドを、一つの前記グループとし、一つのグループとした複数の現パリティ読み出し用コマンドを前記入出力処理装置に渡し、前記パリティ生成手段によって書き込む新パリティが生成された、複数の新パリティ書き込み用コマンドを、一つの前記グループとし、一つのグループとした複数の新パリティ書き込み用コマンドを前記入出力処理装置に渡し、
前記パリティ生成部は、前記入出力処理装置が、渡された現データ読み出し用コマンドに応じて前記ディスクアレイより読み出した前記現データと、新データ書き込み用コマンドによって書き込む新データとの排他的論理和を算出する手段と、算出された現データと新データの排他的論理和を保持する手段と、現パリティ用読み出し用コマンドに応じて前記ディスクアレイより読み出した現パリティと保持したパリティの排他的論理和を、前記新パリティ書き込み用コマンドによって書き込む新パリティとして生成することを特徴とする計算機システム。
【請求項９】請求項８記載の計算機システムであって、
前記ホスト装置のＣＰＵは、前記ディスクアレイより読

み出したデータおよびパリティを、メモリ上に設けたキャッシュ領域に保持する手段を有し、
前記関連付け手段は、前記キャッシュ領域に、現データ読み出し用コマンドによって読み出すべきデータに対応するデータが存在する場合には、現データ読み出し用コマンドを前記入出力処理装置に渡さず、前記キャッシュ領域に、現パリティ読み出し用コマンドによって読み出すべきパリティに対応するパリティが存在する場合には、現パリティ読み出し用コマンドを前記入出力処理装置に渡さず、
前記パリティ生成部は、前記キャッシュ領域に、現データ読み出し用コマンドによって読み出すべきデータに対応する出が存在する場合には、前記ディスクアレイより読み出した前記現データに代えて前記キャッシュ領域に存在するデータを用いて、前記現データと新データとの排他的論理和を生成し、前記キャッシュ領域に、現データ読み出し用コマンドによって読み出すべきパリティに対応するパリティが存在する場合には、前記ディスクアレイより読み出した前記現パリティに代えて前記キャッシュ領域に存在するパリティを用いて、現パリティと保持したパリティの排他的論理和を、前記新パリティ書き込み用コマンドによって書き込む新パリティとして生成することを特徴とする計算機システム。
【請求項１０】ホスト装置と補助記憶装置との間の入出力を担う入出力処理装置であって、
ホスト装置より渡された複数のコマンドに付加されている、当該コマンドの属するグループを示す情報に基づいて、渡された各コマンドと当該コマンドが属するグループの対応を表す情報を作成して記憶する手段と、ホスト装置より渡された各コマンドの指示する補助記憶装置のアクセスを実行する手段と、記憶した各コマンドと当該コマンドが属するグループの対応に基づいて、１つのグループに属する全てのコマンドの指示するアクセスが終了した場合に、当該グループについての処理が終了したことを前記ホスト装置に割り込みを用いて報告する手段とを有することを特徴とする入出力装置装置。
【請求項１１】ホスト装置と補助記憶装置との間の入出力を担う入出力処理装置であって、
前記入出力処理装置は、前記ホスト装置より渡されたコマンドの指示する補助記憶装置のアクセスを実行する手段と、前記ホスト装置に保持されたグループと当該グループに属するコマンドの対応に基づいて、１つのグループに属する全てのコマンドの指示するアクセスが終了した場合に、当該グループについての処理が終了したことを前記ホスト装置に割込みを用いて報告する手段とを有することを特徴とする入出力処理装置。
【請求項１２】割込み処理機構を備えたＣＰＵを備えたホスト装置と、データの記録再生を行なう補助記憶装置と、ホスト装置と補助記憶装置との間の入出力を担う入出力処理装置とを備えた計算機における、補助記憶装置のアクセス方法であって、
前記ホスト装置のＣＰＵにおいて、前記補助記憶装置のアクセスを指示する、複数のコマンドをグループ化するステップと、
前記ホスト装置のＣＰＵにおいて、各コマンドに当該コマンドが属するグループを識別可能な識別情報を付加し、前記識別情報を付加した複数のコマンドを独立に前記入出力処理装置に渡すステップと、
前記入出力処理装置において、渡された複数のコマンドに付加されている前記識別情報より、渡された各コマンドと当該コマンドが属するグループの対応を表す情報を作成して記憶するステップと、
前記入出力処理装置において、前記ＣＰＵより独立に発行された各コマンドの指示する補助記憶装置のアクセスを実行するステップと、
前記入出力処理装置において、記憶した各コマンドと当該コマンドが属するグループの対応に基づいて、１つのグループに属する全てのコマンドの指示するアクセスが終了したかを判定し、終了した場合に、当該グループについての処理が終了したことを前記ＣＰＵに割り込みを用いて報告するステップとを有することを特徴とする補助記憶装置のアクセス方法。
【請求項１３】割込み処理機構を備えたＣＰＵとメモリとを備えたホスト装置と、データの記録再生を行なう補助記憶装置と、ホスト装置と補助記憶装置との間の入出力を担う入出力処理装置とを備えた計算機における、補助記憶装置のアクセス方法であって、
前記ホスト装置のＣＰＵにおいて、前記補助記憶装置のアクセスを指示する、複数のコマンドを、所定の条件に従いグループ化し、前記メモリに、各グループと当該グループに属するコマンドの対応を表す情報を記憶するステップと、
前記複数のコマンドを独立に前記入出力処理装置に渡すステップと、
前記入出力処理装置において、前記ＣＰＵより独立に発行された各コマンドの指示する補助記憶装置のアクセスを実行するステップと、
前記メモリに記憶された各グループと当該グループに属するコマンドの対応に基づいて、１つのグループに属する全てのコマンドの指示するアクセスが終了したかを判定し、終了した場合に、当該グループについての処理が終了したことを前記ＣＰＵに割込みを用いて報告するステップとを有することを特徴とする補助記憶装置のアクセス方法。
【発明の詳細な説明】
【０００１】
【産業上の利用分野】本発明は、主として、計算機システムにおける補助記憶装置へのアクセスの技術に関するものである。
【０００２】

【従来の技術】従来、計算機システムにおいて補助記憶装置として用いられるディスク装置を高速化、高信頼化する技術として、複数台のディスク装置を１台のホスト計算機に接続するマルチディスクと呼ばれる技術が知られている。

【０００３】たとえば、" A Case for Redundant Arrays of Inexpensive Disks(RAID)"InProc. ACM SIGMOD, June 1988（カルフォルニア大学　バークレー校発行）には、複数のディスク装置を専用の制御装置を介してＣＰＵに接続し、複数のディスク装置へのデータの分配制御等のディスク管理をすべてこの制御装置にて実施することにより、ＣＰＵが、複数のディスク装置を１台のディスク装置に見なせるようにしている。

【０００４】また、さらに、近年では、年率２倍弱に達するＣＰＵの高性能化に伴い、専用の制御装置を用いずに、複数のディスク装置の制御をＣＰＵによりソフトウェア的に実施する技術、いわゆるソフトアレイ、またはソフトＲＡＩＤと呼ばれる技術が実現されている。

【０００５】この技術によれば、ＣＰＵのディスクインタフェースに複数台のディスク装置を接続する。そして、ＣＰＵ上で実行されるドライバプロセスが、他のプロセスが、複数台のディスクを仮想的に一台のディスク装置に見なして発行したコマンドを、個々のディスク装置へのコマンドに変換して個々のディスク装置に発行する。また、ドライバプロセスは、記録時にはデータの誤り訂正用の冗長データの生成し、データと共にディスク装置に記録する。また、再生時には、再生データの誤りを、共に記録した誤り訂正用の冗長データを用いて訂正する。

【０００６】このように、ソフトアレイ、またはソフトＲＡＩＤと呼ばれる技術によれば、年々高速化が進むＣＰＵの能力を有効に利用して、低価格で、かつ高性能、高信頼なディスクシステムを実現することができる。

【０００７】ところで、ディスク装置のディスクへのアクセスは、機械的な動作が必要であることから時間がかかる。

【０００８】そこで、従来より、ディスク装置のディスクへのアクセスを、見かけ上、早めるために、従来、リードキャッシュ、ライトキャッシュと呼ばれる技術が広く用いられている。

【０００９】これらの技術では、主記憶上にキャッシュ用の領域を確保する。そして、リードキャッシュによれば、ディスクからデータをリードした際にキャッシュ用の領域に、そのデータを保管し、もし、つぎに同じデータのリード要求がプロセスより発行された際には、ディスクアクセスを行なうこと無く主記憶のキャッシュ用領域からデータをリードし要求元のプロセスに与えることにより、ディスクアクセス回数を削減する。また、ライトキャッシュによれば、ライト時に主記憶上のキャッシュ用の領域にデータを書き込み、ディスクアクセス要求元のプロセスに終了通知を発行する。そして、ある時間後、主記憶上のキャッシュ用の領域のデータをまとめてディスクにライトすることでディスクアクセス回数を削減する。

【００１０】これらのリードキャッシュ、ライトキャッシュの技術は、ディスクアクセス要求元のプロセスの、アクセスの時間的、空間的局所性を利用したもので、効果が大きいため、多くの計算機システムにおいて広く用いられている。

【００１１】

【発明が解決しようとする課題】さて、計算機システムにおいて、ユーザのプログラムは、一般にＯＳ（オペレーティングシステム）という基本プログラムの管理のもとで動作している。ＯＳは、プログラムを並列に実行させる。この並列動作を実現するため、ＯＳはプログラムをプロセスという単位で管理し、これを時分割に切り替えながら動作させている。

【００１２】また、一般的な計算機システムでは、ＣＰＵから、ディスク装置にディスクをアクセスするコマンドが発行され、ディスク装置がアクセスを終了すると、ディスク装置から前記コマンドに対応するアクセスの終了を通知するための割込みがＣＰＵに送られる。ＣＰＵは、この割込みを受け取ると、その時動作しているプロセスを一旦停止し、ディスクアクセス終了のための所定の割込み処理を開始する。

【００１３】ここで、この割込み処理は、次のような手順により行なわれる。すなわち、ＣＰＵは、まず、動作中のプロセスを一旦停止し、このプロセスを再度動作させるために必要な全ての環境を保存し、割り込み処理を実行するのに必要な環境を準備し、所定の割り込み処理を実施し、割り込み処理を実施する以前の環境に戻し、停止したプロセスを再起動する。

【００１４】このように、割込み処理の実行は、ＣＰＵにとって負荷が大きい。

【００１５】一方、前述したソフトアレイ、またはソフトＲＡＩＤと呼ばれる技術によれば、ＣＰＵ上で実行されるドライバプロセスが、個々のディスク装置への複数のコマンドに変換して個々のディスク装置に発行する。したがい、ディスク装置よりの、アクセスの終了を通知するための割込みも、個々のディスクに発行した複数のコマンドのそれぞれについて発生する。

【００１６】また、高信頼化のために誤り訂正用の冗長データを生成し、データと共にディスクに保存する場合には、ＣＰＵは、この冗長データを生成するために必要なデータのリードコマンドや、冗長データのライトコマンドを、ユーザプロセスより要求されたデータのアクセスコマンドに加えて発行するため、さらに、ディスク装置よりの、アクセスの終了を通知するための割込みの発生回数が増加する。

【００１７】すなわち、前述したソフトアレイ、または

ソフトＲＡＩＤと呼ばれる技術によれば、割込み処理の発生頻度が高く、ＣＰＵの、込み処理の実行のための負荷が過大となってしまい、ユーザプログラムを十分なスピードで実行できなくなったり、マルチディスクの能力を充分に引き出せなくなってしまう。

【００１８】一方、前記ライトキャッシュの技術によれば、ある時間毎に、主記憶上のキャッシュ用領域から、まだディスクに書き戻されていないデータを書き出す処理を行う必要がある。また、主記憶の容量には限りがあるため、ディスクに書き戻していないデータ量が一定量を超えたら、次のデータの書き込みにキャッシュ用領域を用いることができるように、ある一定以下の容量に減少するまでデータのディスクへの書き出す処理を行なう必要がある。

【００１９】また、これらの書き出し処理では、どちらも一度にたくさんデータをディスクに書き込む必要があるので、ＣＰＵは、このような書き出し処理のために多くのライトコマンドを、ユーザプロセスより要求されたデータのアクセスコマンドに加えて発行する。このため、ディスク装置よりの、アクセスの終了を通知するための割込みの発生回数が増加する。

【００２０】すなわち、ライトキャッシュの技術でも、割込み処理の発生頻度が高くなるために、ＣＰＵの、込み処理の実行のための負荷が過大となってしなう。

【００２１】そこで、本発明は、ディスクアクセスに起因する、ＣＰＵへの割り込み発生回数を削減することにより、ＣＰＵの負荷を低減することのできる計算機システムを提供することを目的とする。

【００２２】

【課題を解決するための手段】前記目的達成のために本発明は、割込み処理機構を備えたＣＰＵとメモリとを備えたホスト装置と、備えた補助記録媒体にデータの記録再生を行なう補助記憶装置と、ホスト装置と補助記憶装置との間の入出力を担う入出力処理装置とを備えた計算機であって、前記ホスト装置のＣＰＵは、前記補助記憶装置のアクセスを指示する、複数のコマンドを、所定の条件に従いグループ化し、前記メモリに、各グループと当該グループに属するコマンドの対応を表す情報を記憶し、各コマンドに当該コマンドが属するグループを識別可能な識別情報を付加し、前記識別情報を付加した複数のコマンドを独立に前記入出力処理装置に渡す手段と、前記入出力処理装置より特定のグループについての処理が終了したことを割込みによって報告された場合に、前記メモリに記憶した各グループと当該グループに属するコマンドの対応を参照し、前記特定のグループに属する全てのコマンドに関する、所定の終了処理を行なう終了処理手段とを有し、前記入出力処理装置は、渡された複数のコマンドに付加されている前記識別情報より、渡された各コマンドと当該コマンドが属するグループの対応を表す情報を作成して記憶する手段と、前記ＣＰＵより独立に発行された各コマンドの指示する補助記憶装置のアクセスを実行する手段と、記憶した各コマンドと当該コマンドが属するグループの対応に基づいて、１つのグループに属する全てのコマンドの指示するアクセスが終了した場合に、当該グループについての処理が終了したことを前記ＣＰＵに割り込みを用いて報告する手段とを有することを特徴とする計算機システムを提供する。

【００２３】なお、前記入出力処理装置に代えて、前記ＣＰＵより独立に発行された各コマンドの指示する補助記憶装置のアクセスを実行する手段と、前記メモリに記憶された各グループと当該グループに属するコマンドの対応に基づいて、１つのグループに属する全てのコマンドの指示するアクセスが終了した場合に、当該グループについての処理が終了したことを前記ＣＰＵに割込みを用いて報告する手段とを有する入出力処理装置を用いるようにしてもよい。

【００２４】

【作用】本発明に係る計算機システムによれば、個々のコマンドについての処理が終了する毎に前記ＣＰＵに報告する代わりに、各コマンドと当該コマンドが属するグループの対応に基づいて、１つのグループに属する全てのコマンドの指示するアクセスが終了した場合にのみ、当該グループについての処理が終了したことを前記ＣＰＵに割り込みを用いて報告する。したがい、ＣＰＵ割り込み処理の回数を削減でき、ＣＰＵの負荷を軽減できる。

【００２５】

【実施例】以下、本発明に係る計算機システムの実施例について説明する。

【００２６】図１に、本実施例に係る計算機システムの構成を示す。

【００２７】図中、１はホスト計算機（以下、「ホスト」と呼ぶ）、１０１はホスト１の全体制御、各種計算を実行するＣＰＵ、１５１は、ＣＰＵ１０１が実行するプログラムや、データ等を格納するメモリ、１５３はＣＰＵのバスであるローカルバス、１５４は周辺装置等を接続するためのシステムバス、１５２はローカルバス１５３とシステムバス１５４を接続し、ＣＰＵ１０１から周辺回路を接続するためのシステムバスインタフェース（以下、「システムバスＩ／Ｆ」と記す）である。

【００２８】また、２はホスト１のシステムバス１５４に接続するディスク装置の入出力制御を司るＩ／Ｏ処理装置、２０１はＩ／Ｏ処理装置２の全体制御、および各種演算を実行するＭＰＵ、２５１はホスト１のシステムバス１５４に接続するためのシステムバスインタフェース（以下、「システムバスＩ／Ｆ」と記す）、２５２はＩ／Ｏ処理装置の内部バス、２５３はＭＰＵが実行するプログラムや、必要なデータ等を格納するメモリ、２５４はディスクを接続するためのディスクインタフェース（以下、「ディスクＩ／Ｆ」と記す）である。

【００２９】３はディスク装置であり、複数台のハードディスク装置（以下、「ディスク」という）（３０１～３０５）から構成されるディスクアレイである場合もあり、１台もしくは複数台の単体のディスクである場合もある。本第１第実施例では、１台もしくは複数台の単体のディスクで構成されるものとして説明する。

【００３０】ここで、まず、ＣＰＵ１０１のディスクアクセス動作におけるデータの流れを説明する。

【００３１】ホスト１において、ＣＰＵ１０１は、プログラム実行中に、ディスク３からのデータリードが必要になったとすると、システムバスＩ／Ｆ１５２を介してＩ／Ｏ処理装置２にリードコマンドを発行する。Ｉ／Ｏ処理装置２では、システムバスＩ／Ｆ２５１を介して、ＭＰＵ２０１が上記リードコマンドを受信し、これを解析し、対象となるディスク３へのコマンド（ディスクコマンド）へ変換し、ディスクＩ／Ｆ２５４を介し、対象ディスク３にコマンドを発行する。ディスク３はこのディスクコマンドを受信し、ヘッドの位置決め後、対象データをＩ／Ｏ処理装置２に転送する。Ｉ／Ｏ処理装置２はディスクＩ／Ｆ２５４を介してこのデータを受信し、システムバスＩ／Ｆ２５１、およびホストのシステムバス１５４、システムバスＩ／Ｆ１５２経由でホスト１のメモリ１５１に転送する。このデータ転送はホスト１のＣＰＵ１０１が直接Ｉ／Ｏ処理装置のシステムバスＩ／Ｆ２５１からデータをリードし、メモリ１５１に転送する方式と、Ｉ／Ｏ処理装置２が独自にホスト１のメモリ１５１に転送する方式とがあるが、本実施例では、Ｉ／Ｏ処理装置２が独自にホスト１のメモリ１５１に転送するものとする。

【００３２】以上の処理により、ディスク３からのデータのリード処理が完了し、ホスト１のＣＰＵ１０１は、このデータを用い演算を実施できる。なお、ここではリード動作について説明したが、ライト時にはデータ転送方向が逆になるのみで同様である。

【００３３】次に、このような、ＣＰＵ１０１のディスクアクセス動作を実現する、ホスト１のＣＰＵ１０１と、Ｉ／Ｏ処理装置２のＭＰＵ２０１の構成を、コマンド処理系を中心に図２に示す。

【００３４】図中、１０１０はＣＰＵ１０１上で実行されるオペレーティングシステム（ＯＳ）、１０２はディスク上に格納しているファイルのアクセス要求を生成、発行するユーザプロセス、１０２０はキャッシュデータライト要求部、１０００はディスクドライバ部である。

【００３５】なお、ユーザプロセス１０２、キャッシュデータライト要求部１０２０、ディスクドライバ部１０００は、実際には、ＣＰＵ１０１で実行されるプログラムであるオペレーティングシステム１０１０の制御下で、ＣＰＵ１０１で実行されるプロセスとして実現される。

【００３６】なお、図では、ＣＰＵ１０１とＭＰＵ２０１を直接接続して示しているが、実際には、ＣＰＵ１０１とＭＰＵ２０１間の接続は、システムバスＩ／Ｆ１５２、２５１を介して行なわれる。また、ＭＰＵ２０１とディスク装置３を直接接続して示したが、実際には、ＭＰＵ２０１とディスク装置３間の接続は、ディスクＩ／Ｆ２５４を介して行なわれる。

【００３７】ディスクドライバ部１０００中の、１０３はファイルとディスク上のデータ記録アドレスとの対応づけ管理を行い、ユーザプロセス１０２、キャッシュデータライト要求部１０２０により生成、発行されたファイルへのアクセス要求を、実際のディスクへのアクセス要求に変換するファイルアクセス要求管理部、１０４は前記ファイルアクセス要求管理部１０３により発行された、ディスクへのアクセス要求を受信し、各ディスク３へのコマンドを生成し、後述する関連するコマンドの関連付け処理や、コマンドの発行処理、終了処理を行なうディスクコマンド管理部、１０５は前記ディスクコマンド管理部１０４により生成されたディスクコマンドをＩ／Ｏ処理装置２に発行し、またディスクコマンド終了通知を受信するＩ／Ｏ処理装置制御部、１０６は、メモリ１５１上に設けられたキャッシュ用領域であるディスクキャッシュ部である。

【００３８】また、前記ディスクコマンド管理部１０４中の、１０４０は前記ファイルアクセス要求管理部１０３から発行されたディスクアクセス要求を受信するディスクアクセス要求受信手段、１０４１は終了したディスクアクセス要求について前記ファイルアクセス要求管理部に終了報告を発行するディスクアクセス要求終了報告発行手段、１０４２は前記ディスク装置３がディスクアレイである場合に、発行されたアクセス要求を複数台の各ディスクへのアクセス要求に変換するディスクアレイ管理手段、１０４３はディスク装置の管理と、後述するディスクコマンドに関連付けを施すディスクコマンド関連付け管理手段１０４３０を内蔵するディスクコマンド管理手段である。

【００３９】また、１０４４は、前記ディスクコマンド関連付け管理手段１０４３により関連付けられたコマンド群の関連情報を管理する関連コマンド管理テーブル等（後に詳述する）を記憶する、メモリ１５１上に設けたコマンド管理エリアである。

【００４０】次に、ＭＰＵ２０１中、２０１０はＯＳに相当する全体制御部、２０２は前記ホスト１のＩ／Ｏ処理装置制御部１０５が発行したディスクコマンドを受信し、全体制御部２０１０に通知し、ディスクコマンドの終了をホスト１に通知するＩ／Ｏ処理装置Ｉ／Ｆ部、２０３は前記受信したコマンドの登録・終了管理、および、関連付けられたコマンドの登録・終了管理を行うディスクコマンド管理部、２０４は前記ディスクコマンドに従いディスクにアクセスしディスク装置とホスト１のメモリ１５１間のデータ転送を実行するディスクコマン

ド処理部である。なお、Ｉ／Ｏ処理装置Ｉ／Ｆ部２０２、ディスクコマンド管理部２０３、ディスクコマンド処理部２０４は、実際には、ＭＰＵ２０１上で実行されるプログラムである全体制御部２０１０の制御下で実行されるプロセスである。

【００４１】また、２０５は、メモリ２５３上に設けたコマンド管理エリアであり、後述する関連コマンド管理テーブル等を格納する。

【００４２】次に、ディスクコマンド管理部２０３中の、２０３０は前記ホスト１内部のディスクコマンド管理手段１０４３により関連付けられたコマンドの関連付けの内容を、Ｉ／Ｏ処理装置２内のメモリ２５３のコマンド管理エリア２０５に登録する関連コマンド登録手段、２０３１は登録した関連コマンドに対応するディスクアクセスがすべて終了したかどうかを判定する関連コマンド終了判定手段である。

【００４３】以下、本第１実施例におけるディスクアクセス動作について説明する。

【００４４】図２において、実行されているユーザプロセス１０２においてディスク上のデータのリードライトが必要になると、ファイルアクセス要求をＯＳ１０１０に発行する。ＯＳ１０１０は、ディスクドライバ部１０００を起動し、ファイルアクセス要求を渡す。

【００４５】ディスクドライバ部１０００の、ファイルアクセス要求管理部１０３は、内部に備えているファイル格納情報テーブルを参照して、ファイルアクセス要求に従い、アクセスすべきディスク３と、ディスク３上のファイルの転送先頭アドレスと、ディスク３へのデータ転送長を得、これらを含むディスクアクセス要求にファイルアクセス要求を変換し、ディスクコマンド管理部１０４に送出する。

【００４６】ディスクコマンド管理部１０４のディスクアクセス要求受信手段１０４０は、このディスクアクセス要求を受信し、ディスクコマンドパケットをメモリ１５１上に生成する。

【００４７】ディスクコマンドパケットの構成は図３に示すとおりであり、コマンドパケットＣＰには、(1)使用するＩ／Ｏ処理装置（ボード）へのコマンドの種類、(2)対象（ターゲット）ディスクの番号、(3)対象ディスク中の論理装置番号（パーティション番号、論理ユニット番号等）、(4)データ転送長、(5)対象データのディスク上のアドレス、(6)ディスクコマンドの種類（リード／ライト）、(7)データの転送先（元）メモリアドレス、(8)コマンドパケットの実行ステータス情報、(9)ホスト内部またはＩ／Ｏ処理装置内部で複数のコマンドパケットをキューイングする際にキュー管理に用いるコマンドキューイングポインタ（以上(1)から(9)を内容Ａと称する）、(10)関連付けコマンドの管理に必要なコマンド付加パケット管理用のポインタ、等のフィールドを有している。

【００４８】ディスクアクセス要求受信手段１０４０は、このような１つのコマンドパケットの領域をメモリ１５１上に確保し、(1)使用するＩ／Ｏ処理装置（ボード）へのコマンドの種類、(2)対象（ターゲット）ディスクの番号、(3)対象ディスク中の論理装置番号（パーティション番号、論理ユニット番号等）、(4)データ転送長、(5)対象データのディスク上のアドレス、(6)ディスクコマンドの種類（リード／ライト）、(7)データの転送先（元）メモリアドレス等のフィールドに情報を格納した後、ディスクアレイ管理手段１０４２、にそのコマンドパケットのメモリ１５１上のアドレスを渡す。

【００４９】ここで、本第１実施例では、ディスク装置３は単体のディスクより構成されているものとしているので、ディスクアレイ管理手段１０４２は、何もせずに、受け取ったコマンドパケットのアドレスをディスクコマンド管理手段１０４３に渡す。ディスク装置３がディスクアレイ装置である場合のディスクコマンド管理手段１０４３の動作については、後に第２実施例として説明する。

【００５０】ここで、次にコマンドパケットのアドレスを受け取るディスクコマンド管理手段１０４３の構成を図４に示す。

【００５１】図中、１０４３１は、ホスト１に接続されているＩ／Ｏ処理装置２がディスクコマンドの関連付け機能をサポートしているかどうかの判断をするディスクコマンド関連付け機能サポート有無判定手段、１０４３２はディスクコマンドパケットが既に関連付けがなされているかどうかを判定するディスクコマンド関連既所持有無判定手段、１０４３３はコマンドパケットアドレスをキューイングするためのキューイング手段、１０４３４はディスク資源の管理を行うディスク装置管理手段、１０４３０はディスクコマンドに関連付けを施し、また、終了した関連ディスクコマンドの終了処理を行うディスクコマンド関連付け管理手段である。

【００５２】このような構成において、ディスクコマンド管理手段１０４３はコマンドパケットアドレスを受信すると、ディスクコマンド関連付け機能サポート有無判定手段１０４３１は、このメモリ１５１の、このアドレスのコマンドパケットを参照し、コマンドパケットの対象とするディスク装置の接続しているＩ／Ｏ制御装置が、後に説明するような関連付け機能をサポートしているかどうかを判断する。なお、この判断のために、あらかじめ、ホスト１に接続している各Ｉ／Ｏ処理装置２についての情報をディスクコマンド管理部１０４に登録しておくようにする。

【００５３】さて、この判断により、関連付け機能をサポートしていないディスク装置へのコマンドパケットであることが判明したならば、このコマンドパケットのアドレスは直接ディスク装置管理手段１０４３４へ送出される。もし、サポートしているならば、コマンドパケッ

トのアドレスは、次にディスクコマンド関連既所持有無判定手段１０４３２に送出される。

【００５４】ディスクコマンド関連既所持有無判定手段１０４３２は、受け取ったアドレスのディスクコマンドパケットが既に関連付けがなされているかどうかを判定する。ディスクコマンドの関連付けは、後に第２実施例で説明するように、上位に位置するディスクアレイ管理手段１０４０によって関連付けられている場合があるからである。ディスクコマンド関連既所持有無判定手段１０４３２は、既に関連付けが成されている場合は、再度関連付けを施す必要はないため、ディスク装置管理手段１０４３４に、コマンドパケットのアドレスを送出する。

【００５５】一方、ディスクコマンド関連既所持有無判定手段１０４３２は、コマンドパケットに関連付けがまだなされていない場合には、コマンドパケットのアドレスをキューイング手段１０４３３に送出する。キューイング手段１０４３３は、到着したコマンドパケットを、順次、メモリ１５１上に設けたキューに格納する。このとき、コマンドパケットのアドレスが格納されたという情報はディスクコマンド関連付け管理手段１０４３０に通知される。

【００５６】以上で、一つのファイルアクセス要求に対するディスクドライバ部１０００の処理は終了し、ＣＰＵ１０１上では、他のプロセスが実行される。

【００５７】一方、ディスクドライバ部１０００は、ＯＳ１０１０の管理するタイマのタイマ割込みや、ＯＳ１０１０上で動作する所定のディスクコマンド管理プロセス１０２’からの起動命令によって、適当な時間間隔で定期的にも起動される。そして、タイマ割込みによって起動された場合には、以下に説明する処理を実行する。なお、このように、定期的にディスクドライバ部１０００を起動するのではなく、キューイング手段１０４３３より通知された情報より、キューに格納されたコマンドパケットのアドレス数が一定数になった場合に、定期的に起動された場合に行なうものとして説明する以下の処理を、引き続き実行するようにしてもよい。

【００５８】さて、タイマ割込みによって起動された場合、ディスクドライバ部１０００のディスクコマンド関連付け管理手段１０４３０は、キューに格納されているいくつかのコマンドパケットのアドレスの示すメモリ１５１上のコマンドパケットを走査し、関連付け可能なコマンドパケットの群を選びだす。関連付け可能なコマンドパケットの判定基準は、任意に設けることができる。なお、本明細書では、後述する第２実施例以降の実施例でいくつかの具体例を提示する。

【００５９】そして、もし、関連付け可能ないくつかのコマンドパケットを検出したならば、ディスクコマンド関連付け管理手段１０４３０はキューイング手段１０４３３から、順次、これらのコマンドパケットのアドレスを取り出す。そして、これらのコマンドパケットに関連付けを施す。

【００６０】ここで、この関連付けの詳細について説明する。

【００６１】図３は、３つのコマンドパケットＣＰ１、ＣＰ２、ＣＰ３が関連付けの対象として選ばれた場合を示している。いま、この選ばれたコマンドパケットを関連コマンドパケットと呼ぶことにする。また、選ばれたコマンドパケットの集合を関連コマンドパケット群と呼ぶことにする。

【００６２】ディスクコマンド関連付け管理手段１０４３０は、まず、これら各々の関連コマンドパケットに対応するコマンド付加パケットをメモリ１５１上に作成し、コマンドパケット中の関連付けコマンド付加パケットポインタに、コマンド付加パケットのアドレスを格納することによりコマンド付加パケットをコマンドパケットにリンクさせる。

【００６３】ここで、コマンド付加パケットは、(1)関連コマンド群に固有の識別番号（関連コマンドタグ）、(2)関連コマンド群中のコマンドパケットの数、(3)関連コマンドパケットの実行ステータス情報（以上、(1)から(3)を内容Ｂと称する）、(4)対応する関連コマンド管理テーブルへのポインタ、等のフィールドを有している。

【００６４】また、ディスクコマンド関連付け管理手段１０４３０は、一つの関連コマンドパケット群に対して、１つの関連コマンドテーブル１０４４をメモリ１５１上に作成する。関連コマンド管理テーブルは、(1)関連コマンド群の固有の識別番号（関連コマンドタグ）、(2)関連コマンドパケット群中のコマンドパケットの総数、(3)関連コマンドパケット群中の未終了のコマンドパケット数、(4)関連コマンド群の実行ステータス（以上(1)から(4)を内容Ｃと称する）、(5)次に述べるコマンドパケットアドレスポインタテーブルを管理するためのリンクポインタ、等のフィールドを有している。

【００６５】また、ディスクコマンド関連付け管理手段１０４３０は、コマンドパケット毎に、コマンドパケットアドレスポインタテーブルを作成する。コマンドパケットアドレスポインタテーブルには、(1)コマンドパケットへのアドレスポインタ、(2)次のコマンドパケットアドレスポインタテーブルへのポインタ、等のフィールドを設ける。

【００６６】このように、コマンドパケットおよび関連コマンド管理テーブルは、コマンド付加パケット、マンドパケットアドレスポインタテーブルを介して双方向で参照可能な構成となっている。

【００６７】さて、以上のように、ディスクコマンド関連付け管理手段１０４３０は関連付け可能なコマンドパケットを抽出し、メモリ上のコマンド管理エリア１０４４上の関連コマンド管理テーブルに登録する。

【００６８】次にディスクコマンド関連付け管理手段１０４３０はディスク装置管理手段１０４３４にコマンドパケットのアドレスを発行する。

【００６９】ディスク装置管理手段１０４３４は、渡されたアドレスのコマンドパケットに、コマンド付加パケットがリンクされている場合は、このコマンド付加パケットを付加して、Ｉ／Ｏ処理装置制御部１０５に発行する。

【００７０】次に、Ｉ／Ｏ処理装置制御部１０５は、このコマンドパケットを受信し、Ｉ／Ｏ処理装置２に、コマンドパケットを送出する。具体的には、Ｉ／Ｏ処理装置２のシステムＩ／Ｆ部２５１のコマンドレジスタにコマンドパケットを書き込む。

【００７１】以上の処理が終了したらディスクドライバ部１０００は処理を処理し、ＣＰＵ１０１上では、他のプロセスが実行される。

【００７２】一方、Ｉ／Ｏ処理装置２では、システムＩ／Ｆ部２５１のコマンドレジスタに書き込まれたコマンドパケットは、Ｉ／Ｏ処理装置Ｉ／Ｆ部２０２のディスクコマンド受信手段２０２０により、その後メモリ２５３に書き込まれ、その書き込みアドレスが、ディスクコマンド管理部２０３に送出される。

【００７３】ディスクコマンド管理部２０３はこのアドレスを受信し、関連コマンド登録手段２０３０で、このアドレスのコマンドパケットが関連コマンドの１つのコマンドパケットであるか否かを付加されているコマンド付加パケットより判定する。

【００７４】そして、関連付けがなされているならば、メモリ２５３上のコマンド管理エリア２０５の関連コマンドテーブルを参照し、コマンド付加パケットと同じ関連コマンドタグを持った関連コマンド管理テーブルの有無を確認し、もし、存在するならば、コマンド付加パケットの関連コマンド管理テーブルアドレスに、この関連コマンド管理テーブルのアドレスを登録し、さらに、コマンドパケットに対応するコマンドパケットアドレスポインタテーブルを作成し、関連コマンド管理テーブルにリンクする。また、もし、コマンド付加パケットと同じ関連コマンドタグを持った関連コマンド管理テーブルが存在しないならば、新規に関連コマンド管理テーブルを生成し、コマンド付加パケットの関連コマンド管理テーブルアドレスに、この関連コマンド管理テーブルのアドレスを登録し、さらに、コマンドパケットに対応するコマンドパケットアドレスポインタテーブルを作成し、関連コマンド管理テーブルにリンクする。ここで、新規登録時には、関連コマンド管理テーブルの関連コマンドパケット数並びに残コマンド数はコマンド付加パケット中の関連コマンドパケット数を設定しておく。

【００７５】このようにして、ディスクコマンド管理部２０３は、メモリ１５１に作成された図３の情報と、同等の情報をメモリ２５３上に構築する。

【００７６】さて、この後、ディスクコマンド管理部２０３は、メモリ２５３上に設けられている対象ディスクの管理テーブルより、コマンドパケットの対象ディスクの稼動状況を判断し、もし対象ディスクが稼動中であればメモリ２５３上に設けたコマンドキューに、このコマンドパケットのアドレス登録し、ディスクが空き状態になるのを待つ。もし、対象ディスクが空き状態であるならば、対象ディスクの管理テーブルに稼動中であることを設定し、この管理テーブルにコマンドパケットのアドレスを登録する。また、ディスクコマンド処理部２０４に、コマンドパケットのアドレスを渡す。

【００７７】ディスクコマンド処理部２０４は、渡されたアドレスのコマンドパケットに従い、ディスクにアクセスし、ホスト１のメモリ１５１とディスク間のデータ転送を行ない、これが終了したら、終了報告をディスクコマンド管理部２０３に送る。

【００７８】ディスクコマンド管理部２０３の関連コマンド終了判定手段２０３１は、前記ディスク管理テーブルを参照して、終了報告に対応するコマンドパケットを見つけだし、関連コマンドパケットであるかどうかを判定し、もし、関連コマンドパケットあるならば、コマンドパケットに付加されているコマンド付加パケットのフィールド「関連コマンド管理テーブルアドレス」から関連コマンド管理テーブルを参照し、そのフィールド「残コマンド数」の値を１減じた値とする。そして、この引算処理後、残コマンド数が０でないならば（＞０）、同じ関連コマンドパケット群に属する全ての関連コマンドパケットの処理は終了してはいないので、コマンド付加パケットのフィールド「実行ステータス」の値を『コマンド終了待機中』に設定し、何もせずこのまま待機する。もし、引算処理後、残コマンド数が０であるならば、同じ関連コマンドパケット群に属する全ての関連コマンドパケットの処理は終了しているので、ディスクコマンド管理部２０３はＩ／Ｏ処理装置Ｉ／Ｆ部２０２に、終了した関連コマンドパケット群のフィールド「関連コマンドタグ」と、「実行ステータス」を関連コマンド管理テーブルより読み出し通知する。通知を受けた、Ｉ／Ｏ処理装置Ｉ／Ｆ部２０２のディスクコマンド終了報告発行手段２０２１は、ホスト１に、関連コマンドパケット群の処理終了を報告する割込み信号２０６をＣＰＵ１０１に送出する。

【００７９】割り込み信号２０６はＣＰＵ１０１のハードウェアである割込み処理機構を介してＯＳ１０１０に伝えられる。ＯＳ１０１０は、その時、処理中のプロセスを一旦停止し、ＣＰＵ１内部のレジスタ内容等の状態情報（コンテキスト）を保存し、ディスクドライバ部１００１を起動する。

【００８０】割り込み信号２０６に基づき起動された、ディスクドライバ部１００１の、ディスクコマンド終了受信手段１０５１は、Ｉ／Ｏ処理装置２のＩ／Ｏ処理装

置Ｉ／Ｆ部２０２から、関連コマンドタグと、関連コマンドパケット群の実行ステータスを獲得し、関連コマンドパケット群が正常終了したことを確認した後、ディスクコマンド管理手段１０４３にこの獲得した関連コマンドタグと実行ステータスを送出する。

【００８１】ＣＰＵ１０１のディスクドライバ部１００のディスクコマンド管理手段１０４３（図４参照）は、終了した関連コマンドタグを受信し、ディスクコマンド関連付け管理手段１０４３０で、関連コマンドタグにより関連コマンド管理テーブル１０４４を参照し、終了した関連コマンドパケットの適切な終了処理を実施し、関連コマンド群に属する個々のコマンドの終了報告を、ディスクアレイ管理手段１０４２、ディスクアクセス要求終了報告発行手段１０４１、ファイルアクセス要求管理部１０３を介し、ファイルアクセス要求の発行元に終了報告を発行する。また、同時に、関連コマンドタグに対応する関連コマンド管理テーブル１０４４を削除し、当該関連コマンドパケット群に関するすべての処理を終了する。

【００８２】一方、関連コマンド群の実行ステータスをＣＰＵ１０１側に渡したＩ／Ｏ処理装置２のＩ／Ｏ処理装置Ｉ／Ｆ部２０２は、ディスクコマンド管理部２０３に関連コマンドタグを戻し、ディスクコマンド管理部２０３は戻された関連コマンドタグを有する関連コマンド管理テーブル２０５と、これにリンクする関連コマンドパケットコマンドアドレスポインタテーブルと、戻された関連コマンドタグを有する関連コマンドパケットを付加されているコマンド付加パケットと共に削除し、この関連コマンド群に関する全処理を終了する。

【００８３】以上、本第１実施例に係る計算機システムの動作を説明した。このような動作によって、実現されるコマンドと、これに対する報告のシーケンスを図５に示しておく。図５の、横方向は時間の経過を表し、縦軸方向の矢印は、各部位間で渡されるコマンド、報告を表している。

【００８４】図示するように、複数のディスクコマンドを１つの関連コマンド群として管理し、ＣＰＵ１０１とＩ／Ｏ処理装置２での協調動作することで、従来のｎ個のディスクコマンドに対応するｎ個のＩ／Ｏ処理装置からホストへの終了報告を、唯一回の終了通知に削減することができる。

【００８５】さて、前述したように、一般に、この終了報告を行う際には、Ｉ／Ｏ処理装置からホストへ割り込み信号を発行し、ＣＰＵがこれを受信し、現在動作中のプログラム（プロセス）を停止し、割り込み処理を実行するが、この割込み処理にＣＰＵ１０１上のプロセスを切り替えるために通常実行される割込みルーチン、プロセススイッチの処理負荷は大きく、ＣＰＵ１０１の負荷が過大となる。

【００８６】しかし、本第１実施例によれば、割り込み処理を大幅に削減できるので、結局ＣＰＵにかかる負荷率を大幅に低下できる。

【００８７】なお、以上の説明では、Ｉ／Ｏ処理装置２が、ホスト１よりコマンド付加パケットが付加されたコマンドパケットを受け取り、これに基づいて、Ｉ／Ｏ処理装置２内のメモリ２５３に関連コマンド管理テーブル、コマンドパケットアドレスポインタテーブルを作成したが、Ｉ／Ｏ処理装置２が、ホスト１よりコマンドパケットのアドレスにみを受け取り、以降の処理は、このアドレスに基づいて、ホスト１内のメモリ１５１内の、コマンドパケット、コマンド付加パケット、関連コマンド管理テーブル、コマンドパケットアドレスポインタテーブルに直接アクセスして行なうようにしてもよい。すなわち、メモリ２５３にコマンド付加パケットが付加されたコマンドパケット、関連コマンド管理テーブル、コマンドパケットアドレスポインタテーブルを記憶し、これを用いる代わりに、直接、ホスト１内のメモリ１５１内の、コマンドパケット、コマンド付加パケット、関連コマンド管理テーブル、コマンドパケットアドレスポインタテーブルを用いるようにしてもよい。

【００８８】以下、本発明の第２の実施例について説明する。

【００８９】本第２実施例は、第１実施例に係る計算機システムおけるディスク装置３がディスクアレイである場合についてのものである。

【００９０】ディスクアレイは、複数台のディスクから構成され、これらのディスクにデータを分散し格納するものであるが、ユーザプロセスからは仮想的に１台のディスクに見えるようにする必要がある。ユーザプロセスから仮想的に１台のディスクに見えるようにするための処理は、ホストＣＰＵ１０１が行なう。したがい、ディスクへコマンドを発行するＩ／Ｏ処理装置２はディスクアレイのデータ分散管理には一切関与しない。

【００９１】本第２実施例に係る計算機システムの構成は、図１、２、４に示した前記第１実施例に係る計算機システムの構成と同一である。

【００９２】ここで、前記第１実施例で説明を行なわなかったディスクアレイ管理手段１０４２（図２参照）の構成を図６に示す。

【００９３】図６中、１０４２１はｎ台のディスクにデータを分配するためのアドレス変換作業を実施するディスクアレイアドレス変換手段、１０４２２はディスクアレイの構成情報を格納管理するディスクアレイ情報管理手段、１０４２３はディスクアレイ中の任意の１台のディスク装置が故障してもデータを損失しないようにＥＣＣ（エラー訂正コード）を生成するＥＣＣ演算手段、１０４２４はディスクアレイへのコマンド群に関連付けを行うディスクアレイコマンド関連付け管理手段、１０４２５はＥＣＣ生成を高速化するためのＥＣＣおよびその生成に必要なデータを保持しておくＥＣＣ生成キャッシ

ュ手段である。

【００９４】以下、本第２実施例に係る計算機システムの動作について説明する。

【００９５】ただし、本第２実施例では、前記ＥＣＣ演算手段１０４２３および前記ＥＣＣ生成キャッシュ手段１０４２５を用いない場合について説明する。ＥＣＣ演算手段１０４２３を用いる場合については第３実施例として、ＥＣＣ生成キャッシュ手段１０４２５を用いる場合については第５実施例として、後に説明する。

【００９６】ディスクアクセス要求受信手段１０４０は、ユーザプロセス１０２またはディスクキャッシュ管理部１０６よりフィルアクセス要求を受け取ると、ディスクアレイ全体に対するディスクアクセス要求について親コマンドパケットを生成し、メモリ１５１に格納し、その格納アドレスをディスクアレイ管理手段１０４２に渡す。

【００９７】ディスクアレイ管理手段１０４２のディスクアレイアドレス変換手段１０４２１（図６参照）は、この親コマンドパケット受け取り、ディスクアレイ情報管理手段１０４２２の構成情報を参照して、受け取った親コマンドパケットより、個々のディスクへの子コマンドパケット群を生成し、メモリ１５１に格納する。

【００９８】たとえば、ディスクアレイ全体への１つのリードアクセス要求を示す親コマンドパケットが発行されたときに、図７に示すように、そのデータ転送長がｉバイトであり、ディスクアレイにおけるデータ分配単位（ストライプサイズ）がｋバイトで、４ｋ＞ｉ＞３ｋの関係にあれば、要求されたデータは４台のディスクに分散していることになるので、ディスク１～４の４台への４つの子コマンドパケットを生成する。

【００９９】なお、子コマンドパケット群の生成の際には、親、子コマンドパケットに、(1)パケットの階層を示すパケット種、(2)分割コマンド数、(3)子コマンドパケットへのポインタ、(4)自分と同じ階層の次のコマンドパケットへのポインタ、(5)親コマンドパケットへのポインタ等のフィールドを付加し、コマンドパケットを階層的に管理することができるようにする。

【０１００】さて、ディスクアレイアドレス変換手段１０４２１は、子コマンドパケット群を生成しメモリ１５１に格納すると、これらのアドレスをディスクアレイコマンド関連付け管理手段１０４２４に送出する。

【０１０１】ディスクアレイコマンド関連付け管理手段１０４２４は、これらの子コマンドパケット群のアドレスを受信し、子コマンドパケット群に属する子コマンドパケット間に関連付けを施す。関連付けは、前記第１実施例において説明したディスクコマンド関連付け管理手段１０４３０の行なう関連付けと同じであり、メモリ１５１上にコマンド付加パケット、関連コマンド管理テーブル１０４４、コマンドパケットアドレスポインタテーブルを図９に示すとおりに作成することにより行なわれる。すなわち、子コマンドパケットは、前記第１実施例における関連コマンドパケットとなる。

【０１０２】この後、ディスクアレイ管理手段１０４２はディスクコマンド管理手段１０４３に各子コマンドパケットのアドレスを渡す。

【０１０３】ディスクコマンド管理手段１０４２は、子コマンドパケットのアドレスを受信するが、これらの子コマンドパケット間には、既に関連付けが実施されているので、前記第１実施例で述べたように、ディスク装置管理手段１０４３４での関連付けは行わず、ディスク装置管理手段１０４３４に渡し、ディスク装置管理手段１０４３４に渡されたアドレスのコマンドパケットに、これにリンクしているコマンド付加パケットを付加して、Ｉ／Ｏ処理装置制御部１０５よりＩ／Ｏ処理装置に発行する。

【０１０４】その後の、Ｉ／Ｏ処理装置２、ディスク装置３の動作は前記第１実施例で説明した動作と同じであり、Ｉ／Ｏ処理装置２、ディスク装置３は、親コマンドパケットを意識することなく動作する。

【０１０５】さて、前記第１実施例と同様にしてＩ／Ｏ処理装置２よりの関連コマンドパケット群の終了報告を受けたディスクコマンド管理手段１０４３は、終了した関連コマンドタグを受信し、ディスクコマンド関連付け管理手段１０４３０で、関連コマンドタグにより、終了した関連コマンドパケットを見つけ、親コマンドパケットへのポインタの存在を判定し、存在していれば、終了報告と関連コマンドタグをディスクアレイ管理手段１０４２内部のディスクアレイコマンド関連付け管理手段１０４２４に通知する。ディスクアレイコマンド関連付け管理手段１０４２４は、関連コマンドタグにより、終了した関連コマンドパケットを見つけ、親コマンドパケットへのポインタの示す親コマンドパケットに順次リンクしている子コマンドパケット、付加コマンドパケット、関連コマンド管理テーブル、コマンドパケットアドレスポインタテーブルをメモリ１５１より削除すると共に、親コマンドパケットのフィールド「実行ステータス」に正常終了ステータスを登録後、ディスクアクセス要求終了発行手段１０４１を介しファイルアクセス要求管理部１０３に終了報告を行なう。

【０１０６】このように、本第２実施例によれば、ディスクアレイを用いることで従来の単体ディスクへのアクセスに対し大幅に増加するＣＰＵへの割り込み処理に起因するＣＰＵ負荷の上昇を従来の単体ディスク同等に低下させることができる。

【０１０７】本第３実施例は、前記第２実施例に係る計算機システムおけるディスクアレイが、ＥＣＣ機能を有するディスクアレイ、すなわち、ＲＡＩＤ（Ｒｅｄｕｎｄａｎｔ　Ａｒｒａｙｓ　ｏｆ　Ｉｎｅｘｐｅｎｓｉｖｅ　Ｄｉｓｋｓ）である場合、すなわち、前記第１実施例におけるディスク装置３がＲＡＩＤである場合につい

てのものである。

【０１０８】なお、ＲＡＩＤの技術分野ではＥＣＣデータのことを一般にパリティと呼ぶので、以下の説明でも、その名称を用いることにする。

【０１０９】図８に、本第３実施例に係るディスク装置３の構成例を示す。

【０１１０】図８に示した例は、ディスク数が５台のＲＡＩＤについてのものである。５台のディスクのうち、１台はパリティディスクと呼ばれ、ＥＣＣデータ専用のディスクである。ディスク０からディスク３までのデータＤ０，Ｄ１、Ｄ２、Ｄ３の排他的論理和により得られる、パリティデータＰが、ディスク４に格納される。すなわち、Ｐ＝Ｄ０＋Ｄ１＋Ｄ２＋Ｄ３（ただし、本明細書中では、＋は排他的論理和を表すものとする）である。

【０１１１】ここで、ディスク０から４までのディスクの同じディスク内アドレスの領域の集合であるストライプのデータ群をパリティグループと呼ぶ。

【０１１２】このようなＲＡＩＤでは、ディスク１のデータＤｉを変更するならば、対応するパリティデータＰｉも変更する必要がある。この新しいパリティデータＰｉの算出は、新たなデータＤｉ［ｎｅｗ］（新データ）を書き込む以前に、既にディスクに書き込まれているデータＤｉ［ｏｌｄ］（旧データ）とパリティデータＰ［ｏｌｄ］（旧パリティ）を読みだし、次の演算により新しいパリティデータＰ［ｎｅｗ］（新パリティ）を生成することにより行なわれる。

【０１１３】Ｐ［ｎｅｗ］＝Ｄ［ｎｅｗ］＋Ｄ［ｏｌｄ］＋Ｐ［ｏｌｄ］

以上のように、ＲＡＩＤでは、ディスクアレイのうちの１台のディスクへのライト要求であっても、パリティデータ更新のためのディスクアクセスが発生するので、計４回のディスクアクセスを行なう必要がある。

【０１１４】したがい、ＲＡＩＤを用いる場合、ＣＰＵ１０１の負荷は、普通の１台のディスクを用いる場合に対し、割り込み処理に要する負荷が４倍に増加することになる。

【０１１５】以下、本実施例に係る計算機システムの動作について説明する。

【０１１６】本第３実施例に係る計算機システムの構成は、前記第２実施例に係る計算機システムと同じである。

【０１１７】さて、本第３実施例では、ディスクアレイ管理手段１０４２（図２参照）のディスクアレイアドレス変換手段１０４２１（図６参照）は、ディスクアレイへのデータＤｉのライトを要求する一つのコマンドパケットのアドレスを受信すると、まず、コマンドパケットを参照する。そして、ディスクアレイ情報管理手段１０４２２の構成情報を参照して、ディスクの台数、パリティディスクの位置、ストライプサイズの大きさ等を判断し、ライトコマンドパケットにより書き込みを要求されたデータを書き込むべきデータの位置はディスク１のＤｉであることを決定する。

【０１１８】次に同様にして、パリティの格納ディスクおよび格納位置を特定する。

【０１１９】そして、図８のディスク１のデータＤｉへのリード要求と、ライト要求と、ディスク４のＰｉのリード要求と、ライト要求についての４つのコマンドパケットを生成し、メモリ１５１に書き込む。なお、この際、旧データおよび旧パリティをロードするメモリ１５１上の領域、および、新パリティを生成し格納するメモリ１５１の領域を確保し、そのアドレスを、それぞれのコマンドパケットに登録する。

【０１２０】次に、ディスクアレイアドレス変換手段１０４２１は、これらのコマンドパケットのアドレスをディスクアレイ関連付け管理手段１０４２４に渡す。

【０１２１】ディスクアレイ関連付け管理手段１０４２４は、受け取ったコマンドパケット群に対し関連付けを施す。ここで、関連付けは、次のように行なう。

【０１２２】旧データのリードおよび、旧パリティのリード用のコマンドパケットは、メモリ１５１上にコマンド付加パケット、関連コマンド管理テーブル１０４４、コマンドパケットアドレスポインタテーブルを、前記第２実施例と同様に作成することにより関連付ける。これは、両方とも完了しない限り新パリティは生成できないので、両者が終了した時点でＩ／Ｏ処理装置２から終了報告を受ければよいからである。

【０１２３】また、生成した新パリティのライトと、新データのライトのコマンドパケットも、メモリ１５１上にコマンド付加パケット、関連コマンド管理テーブル１０４４、コマンドパケットアドレスポインタテーブルを、前記第２実施例と同様に作成することにより関連付ける。これは両方のライト終了後に、上位にあるファイルアクセス要求手段１０３へライトアクセス要求の終了報告を行なえばよいので、両者が終了した時点でＩ／Ｏ処理装置２から終了報告を受ければ足りるからである。

【０１２４】さて、本第３実施例では、次のようにして、ファイルアクセス要求手段１０３からのディスクアクセス要求に相当するコマンドパケットと、これに基づき作成された２つのリード用のコマンドパケットと２つのライト用のコマンドパケットを関連付ける。

【０１２５】すなわち、図１０に示すように、第２実施例より、階層を１つ増やし、３階層で各コマンドパケット管理するようにする。

【０１２６】第１の階層（親コマンドパケット）は、ファイルアクセス要求手段１０３からのディスクアクセス要求に対応するものである。第２の階層（子コマンドパケット）は、第１の階層のコマンドパケットがライトについてのものである場合に設けるもので、リード処理に対応する子リードコマンドパケットと、ライト処理に対

応する子ライトコマンドパケットが作成される。図示するように、両パケットは親コマンドパケットにリンクしている。

【０１２７】第３の階層（孫コマンドパケット）は、Ｉ／Ｏ処理装置２が処理する各々のディスクコマンドをあらわしたものである。２つのリード用孫コマンドパケット（旧データリードコマンドパケット、旧パリティデータリードコマンドパケット）は子リードコマンドパケットにリンクし２つのライト用孫コマンドパケット（新データライトコマンドパケット、新パリティデータライトコマンドパケット）は子ライトコマンドパケットにリンクしている。

【０１２８】以上のとおり、関連付けが終了したら、ディスクアレイ関連付け管理手段１０４２４は、まず、子リードコマンドパケットにリンクする２つの孫コマンドパケットをディスクコマンド管理手段１０４３に送出する。

【０１２９】また、この後、ディスクコマンド管理手段１０４３は、受信したコマンドパケットのリンクする第１層の親コマンドパケットがライトについてのものである場合には、全てのディスクの、ライトの対象となるストライプに属する領域のロックを設定し、ロックが解除されるまで、受信したコマンドパケットのリンクする第１層の親コマンドパケットにリンクする孫コマンドパケット以外のコマンドパケットの処理を行なわないようにする（ディスク資源のロック）。これは、ライト時の旧（パリティ）データのリードと新（パリティ）データのライトは必ず連続して処理を行う必要があるためである。

【０１３０】ディスクコマンド管理手段１０４３は、この後、これらの孫コマンドパケットにリンクする付加コマンドパケットを付加してディスクコマンド発行手段１０５０に渡し、ディスクコマンド発行手段１０５０該部１０５０は、これをＩ／Ｏ処理装置２に発行する。

【０１３１】Ｉ／Ｏ処理装置２では、前記第１実施例と同様に、２つのリード用コマンドパケットを処理し、両者が終了すると、ホスト１のＯＳ１０１０に一度の終了割り込を発行する。前記第１実施例と同様に、これにより起動されるディスクコマンド終了受信手段１０５１はＩ／Ｏ処理装置２から終了した関連コマンドの関連コマンドタグを獲得し、ディスクコマンド管理手段１０４３に渡す。ディスクコマンド管理手段１０４３は、前記第２実施例と同様に、これをディスクアレイ管理手段１０４２に渡す。

【０１３２】さて、ディスクアレイ管理手段１０４２のディスクアレイ関連付け管理手段１０４２４は、受け取った関連コマンドタグより、処理が終了した子パケットを判定する。そして、子リードパケットの処理が終了したことを確認したならば、子リードパケットの実行ステータスに正常終了を登録し、ＥＣＣ演算手段１０４２３に処理を渡す。

【０１３３】ＥＣＣ演算手段１０４２３はこれを受け、処理が終了した子リードパケットに親コマンドパケット、子ライトコマンドパケットを介してリンクしているライト用孫コマンドパケットのうちの書き込みデータについてのものから新データの格納されているメモリアドレスと、処理が終了した子リードパケットにリンクしている２つのリード用孫コマンドパケットから旧データ、旧パリティデータの格納されているメモリアドレスを獲得する。

【０１３４】そして、これら３データをメモリ１５１より読み出し、排他的論理和を演算し、メモリ１５１の、新パリティライト用の孫コマンドパケットに記述されている格納アドレスに、排他的論理和演算結果を格納する。

【０１３５】ＥＣＣ演算手段１０４２３は、この処理のあと、処理を再びディスクアレイ管理手段１０４２に戻す。ディスクアレイ管理手段１０４２は新パリティの生成が終了したので、子ライトコマンドパケットにリンクする２つのライト用孫コマンドパケットのアドレスをディスクコマンド管理手段１０４３に送出する。

【０１３６】ディスクコマンド管理手段１０４３は、ディスクコマンド発行手段１０５０を介して、Ｉ／Ｏ処理装置２に２つのライト用孫コマンドパケットを発行する。

【０１３７】Ｉ／Ｏ処理装置２では、前記第１実施例と同様に、２つのライト用コマンドパケットを処理し、両者が終了すると、ホスト１のＯＳ１０１０に一度の終了割り込を発行する。前記第１実施例と同様に、これにより起動されるディスクコマンド終了受信手段１０５１はＩ／Ｏ処理装置２から終了した関連コマンドの関連コマンドタグを獲得し、ディスクアレイ管理手段１０４２に渡す。

【０１３８】ディスクコマンド管理手段１０４３は、前記第２実施例と同様に、これをディスクアレイ管理手段１０４２に渡す。また、ディスクコマンド管理手段１０４３は、受け取った関連コマンドタグに対応するライト用孫コマンドパケットが、ライトを要求する第１層の親コマンドパケットにリンクするものである場合には、この親コマンドについて行なったロックの設定を解除する。

【０１３９】一方、ディスクアレイ管理手段１０４２のディスクアレイ関連付け管理手段１０４２４は、受け取った関連コマンドタグより、処理が終了した子パケットを判定する。そして、子リードパケットの処理が終了したことを確認したならば、子ライトコマンドパケットの実行ステータスに正常終了を登録し、子ライトコマンドパケットにリンクする親コマンドパケットにリンクする子リードコマンドパケットの実行ステータスが正常終了であることを確認し、親コマンドパケットの実行ステー

タスに正常終了ステータスを登録後、ディスクアクセス要求終了発行手段１０４１を介しファイルアクセス要求管理部１０３に終了報告をし、すべての子コマンドパケット、孫コマンドパケット、コマンド付加パケット、関連コマンド管理テーブル、コマンドパケットアドレスポインタテーブルを削除する。

【０１４０】一方、ファイルアクセス要求管理部１０３は、親コマンドパケットの正常終了ステータスを確認後、ファイルアクセス要求元に正常終了報告を行ない、親コマンドパケットをメモリ１５１より削除する。

【０１４１】以上のように、本第３実施例によればＲＡＩＤをディスク装置として用いる場合に、ライト時に行なわれる複数のディスクアクセスに関連付けを施し、ＣＰＵ割り込み処理の増加を防ぐことができる。

【０１４２】以下、本発明の第４の実施例について説明する。

【０１４３】本第４実施例は、前記第１実施例においてディスクキャッシュ部１０６を使用した場合についてのものである。

【０１４４】本第４実施例に係る計算機システムの構成を前記第１実施例に係る計算機システムの構成と同じである。

【０１４５】本第４実施例では、ファイル要求管理部１０６は、ディスクからデータをリードした際にディスクキャッシュ部１０６に、そのデータを保管し、もし、つぎに同じデータのリード要求がユーザプロセス１０２より発行された際には、ディスクアクセスを行なうこと無く主記憶のキャッシュ用領域からデータをリードし要求元のユーザプロセスに与える。また、ライト時にディスクキャッシュ部１０６にデータを書き込み、ライト要求元のユーザプロセス１０２に終了通知を発行する。また、ファイル要求管理部１０６には、使用状況を管理する管理テーブルを設ける。

【０１４６】さて、このようにして、ユーザプロセスよりのライト要求に応じてディスクキャッシュ部１０６に書き込んだデータは、ある時間後、まとめてディスク装置３にライトする必要がある。

【０１４７】以下、このディスクキャッシュ部１０６のデータのディスク装置３への書き出し動作について説明する。

【０１４８】いま、図１１に示すように、メモリ１５１上のディスクキャッシュ部１０６には、Ａ、Ｂ、Ｃ、Ｄ、Ｅのデータブロックがディスクに書き込まれていない状態（これをダーティブロックと呼ぶ）で保持されている。このダーティブロックは適当な時間間隔でディスクに書出される。この書き出しのタイミングの決定の方式としては、いくつかの方式が知られているが、本第４実施例では、一定時間毎に検査し、もしダーティブロックが存在するならばこれらを全て書き出すものとして説明する。

【０１４９】さて、ディレイドライトキャッシュデータライト要求部１０２０は一定時間毎にダーティキャッシュ書きだし命令をファイルアクセス要求管理部１０３に送出する。ファイルアクセス要求管理部１０３はこれを受信し、ディスクキャッシュ部１０６の使用状況を管理する管理テーブルを操作し、ダーティブロックを検出する。

【０１５０】ファイルアクセス要求管理部１０３は、これら検出したダーティブロックのディスクライトアクセス要求を作成し、順次、ディスクコマンド管理部１０４に渡す。

【０１５１】ディスクアクセス要求手段１０４０は、これらのディスクアクセス要求を受信し、対応するコマンドパケットを生成してメモリ１５１上に書き込み、そのアドレスを、ディスクコマンド管理手段１０４３に送出する。ディスクコマンド管理手段１０４３では、渡されたアドレスのコマンドパケットを読み出し、前記第１実施例と同様に、ディスクコマンド関連付け機能サポート有無判定手段１０４３１でＩ／Ｏ処理装置２の機能を確認し、ディスクコマンド関連既所持有無判定手段１０４３２で関連既所持のないことを判定し、キューイング手段１０４３３に、そのアドレスを格納する。このようなダーティブロックの書き出しは一度に多数のコマンドパケットを同時に発生させるので、キューイング手段１０４３３には多数のコマンドパケットのアドレスが格納される。

【０１５２】ディスクコマンド関連付け管理手段１０４３０はキューイング手段１０４３３を検査し、アドレスが存在する全てのライトコマンドパケットに、またはある規定個数のライトコマンドパケットに、または、複数のディスクがＩ／Ｏ処理装置２に接続している場合には各ディスクにつき１つずつ抽出したコマンドパケットに関連付けを施す。

【０１５３】ここで、ライトキャッシュを行なっているディスクキャッシュを持つホストでは、ユーザのライト要求はディスクキャッシュ部１０６に書き込んだ時点で終了しており、関連付けにより一つ一つのダーティブロックライトの書きだし処理終了が若干遅延しても問題はほとんど発生しないので、このように多数のライトコマンドパケットに関連付けを施しても問題は生じない。しかし、まれにキャッシュあふれ等により即時に処理を実施したいという要求がある場合には、ファイルアクセス要求管理部１０３からの指定に応じて、コマンドパケットに関連付けを釣支するフィールドを設け、このようなコマンドパケットについては関連付けの対象としないようにしてもよい。

【０１５４】以下の動作は、前記第１実施例と同じである。

【０１５５】なお、本実施例は、前記第２、第３実施例と組み合わせて適用することもできる。すなわち、第

２、第３実施例のディスクアレイ管理手段１０４２において、複数の親コマンドパケットにリンクする、複数のリード用系コマンドパケット間、および、複数のライト用系コマンドパケット間で関連付けを施すようにしてもよい。

【０１５６】以上のように、本第４実施例によれば、ライトキャッシュを有するホスト１において、ＣＰＵの負荷を低減することができる。

【０１５７】以下、本発明の第５実施例について説明する。

【０１５８】本第５実施例は、前記第３実施例において、ディスクアレイ管理手段１０４２のＥＣＣ（パリティ）生成キャッシュ手段１０４２５を用いる場合についてのものである。

【０１５９】前記第３実施例では新パリティを生成する際に、すべての必要なデータ（旧データ、旧パリティ）をディスクから読みだした後に、新パリティを計算し、パリティディスクに新データと共に送出していた。しかし、これでは、１回のライトにつき必ず２台のディスクを占有することになる。

【０１６０】そこで、本第５実施例では、次のようにして、一度に占有するディスクを１台とすることにより、並列度を高める。

【０１６１】すなわち、まず、旧データをリードして新データをライトし、新旧データの排他的論理和を計算し、両データの差分データまでを生成し、メモリ１５１に保存する。そして、適当な遅延時間後に旧パリティをリードして旧パリティと上記差分データとの排他的論理和を計算し新パリテイ生成してライトする。

【０１６２】また、本第５実施例では、より高速化するために、旧データを毎回読みだすのではなく、前記ＥＣＣ生成キャッシュ手段１０４２５が管理するメモリ１５１上の所定の領域に、リードしたデータを保持しておく。このようにすることにより、旧データとして、このデータを用いることができる可能性がある。この場合、旧データを読みだす必要がなくなる。また、さらに、旧パリティも毎回読み出すのではなく、前記ＥＣＣ生成キャッシュ手段１０４２５が管理するメモリ１５１上の所定の領域に、リードしたパリティを保持しておく。これにより、旧パリティとして、このデータを用いることができる可能性がある。また、ディスクに書き込む新データや新パリティを保持しておく。これにより、その後、旧データ、旧パリティとして、これらを用いることができる可能性がある。

【０１６３】なお、このようなデータやパリティの保持は前記第４実施例で示したディスクキャッシュ部１０６で行なうようにしてもよい。

【０１６４】さて、ファイルアクセス要求管理部１０３より、ライトアクセス要求のコマンドパケットを受け取ると、ディスクアドレス変換手段１０４２１は、図１２に示すように、これを親コマンドパケット＃１とし、これをコピーすることによりメモリ１５１上に作成した親コマンドパケットを＃２とする。また、図示するように、親コマンドパケットに、ＮＥＸＴという名称のフィールドを設け、これを用いて、親コマンドパケット＃１と親コマンドパケット＃２をリンクする。

【０１６５】そして、次に、親コマンドパケット＃１の子コマンドパケットとして、旧データをリードするための子コマンドパケット＃１と、旧データをライトするためのコマンドパケット＃２を生成し、親コマンドパケット＃２の子コマンドパケットとして、旧パリティをリードするための子コマンドパケット＃３と、旧パリティをライトするためのコマンドパケット＃４を生成し、４つの子コマンドパケットを生成してメモリ１５１に格納し、これらのアドレスを、ディスクアレイ関連付け管理手段１０４２４に渡す。

【０１６６】ディスクアレイ関連付け管理手段１０４２４は、まず、旧データをリードするためのコマンドパケット＃１を参照し、この旧データをＥＣＣ生成キャッシュ手段１０４２５管理しているメモリ１５１上の領域に存在しているかを判定し、保持していれば、この子コマンドパケット＃１の実行ステータスを正常終了とし、終了報告をディスクアドレス変換手段１０４２１に行なう。ＥＣＣ演算手段１０４２３では、旧データをリードするためのコマンドパケット＃１に親パケット＃１、＃２を介してリンクしている新パリティライト用のコマンドパケット＃４を参照し、フィールド「メモリアドレス」にアドレスが登録されていない場合には、旧データと、新データライト用のコマンドパケットが指定する新データとの排他的論理和（差分データ）を計算して、ＥＣＣキャッシュ生成手段１０４２５を介してメモリ１５１に格納し、そのメモリアドレスを新パリティライト用のコマンドパケット＃４のフィールド「メモリアドレス」に登録する。

【０１６７】もし、旧データを保持していなければ、この子コマンドパケット＃１はディスクコマンド管理手段１０４３に渡され、ディスクに送出し、旧データを読みだし、処理が終了した旧データリード用のコマンドパケットの実行ステータスを正常終了とし、ディスクアドレス変換手段１０４２１に終了報告を行なう。ＥＣＣ演算手段１０４２３は、先程の場合と同様に、旧データをリードするためのコマンドパケット＃１に親パケット＃１、＃２を介してリンクしている新パリティライト用のコマンドパケット＃４を参照し、フィールド「メモリアドレス」にアドレスが登録されていない場合には、旧データと、新データライト用のコマンドパケットが指定する新データの排他的論理和を計算して、ＥＣＣ生成キャッシュ手段１０４２５を介してメモリ１５１に格納し、そのメモリアドレスを新パリティライト用のコマンドパケット＃４のフィールド「メモリアドレス」に登録す

る。

【０１６８】次に、ディスクアレイ関連付け管理手段１０４２４は、新データライト用のコマンドパケット＃２をディスクコマンド管理手段１０４３に渡する。ディスクコマンド管理手段１０４３は、前記第４実施例と同様、他のライトコマンドパケットとの間で関連付けを施された後、ディスクに送出される。以降、前記第１実施例と同様に処理が行なわれ、すべての関連コマンドパケットに対応するデータ転送が終了したならば１回の割り込みがＩ／Ｏ処理装置から発行され、ディスクアレイコマンド関連付け管理手段１０４２４に処理が戻り、ディスクアレイコマンド関連付け管理手段１０４２４は、処理が終了した新データライト用のコマンドパケットの実行ステータスを正常終了とし、ディスクアドレス変換手段１０４２１に終了報告を行なう。ディスクアドレス変換手段１０４２１は、これより親コマンドパケット＃１の終了を認識し、親コマンドパケット＃１の実行ステータスを正常終了とし、ファイルアクセス管理部１０３に終了報告を行う。ファイルアクセス管理部１０３は、ライトアクセスの要求元のユーザプロセス１０２に終了報告を行なう。

【０１６９】なお、実際には、この時点では、パリティの更新が終了していないので、ユーザプロセス１０２への終了方報告は、パリティの更新に先行して行なわれることになる。

【０１７０】次に、ディスクアレイ関連付け管理手段１０４２４は、旧パリティリードのための子コマンドパケット＃３を参照して、旧パリティがＥＣＣ生成キャッシュ手段１０４２５が管理するメモリ１５１上の領域に存在しているかを判定し、存在すれば、このコマンドパケットの実行ステータスを正常終了とし、終了報告をディスクアドレス変換手段１０４２１に行なう。また、ＥＣＣ演算手段１０４２３は、旧パリティリードのための子コマンドパケット＃３に親コマンドパケット＃２を介してリンクしている新パリティのライトのための子コマンドパケット＃４のフィールド「メモリアドレス」に登録されたアドレスに登録されている旧データと新データの差分データをＥＣＣ生成キャッシュ手段１０４２５を介して読み出し、これと、獲得した旧パリティの排他的論理和を計算し、子コマンドパケット＃４のフィールド「メモリアドレス」に登録されたメモリ１５１のアドレスに書き込む。そして、ＥＣＣ生成キャッシュ手段１０４２５内部のキューイング手段に、この子コマンドパケット＃４をキューイングする。

【０１７１】また、もし、旧パリティをＥＣＣ生成キャッシュ手段１０４２５が保持していない場合は旧パリティ用、新パリティ用の両方の子コマンドパケット＃３、＃４を前記ＥＣＣ生成キャッシュ手段１０４２５のキューイング手段にキューイングする。

【０１７２】さて、適当な時間経過後、ＥＣＣ生成キャッシュ手段１０４２５のキューイング手段内のコマンドパケットの発行を行なう。すなわち、ディスクアレイコマンド関連付け管理手段１０４２４は、ＥＣＣ生成キャッシュ手段１０４２５のキューイング手段を走査し、適当な個数の旧パリティリード用子コマンドパケット＃３のみに関連付けを施し、子コマンドパケット＃３をディスクコマンド管理手段１０４３に送出する。以降、前記第１実施例と同様に処理が行なわれ、すべての関連コマンドパケットに対応するデータ転送が終了したならば上記第１、２、３、４の実施例同様１回の割り込みがＩ／Ｏ処理装置から発行され、ディスクアレイコマンド関連付け管理手段１０４２４に処理が戻り、ディスクアレイコマンド関連付け管理手段１０４２４は、処理が終了した旧パリティリード用のコマンドパケットの実行ステータスを正常終了とし、ディスクアドレス変換手段１０４２１に終了報告を行なう。

【０１７３】さて、このようにして、旧パリティ用のリードコマンドパケット＃３が正常終了して旧パリティが得られたら、ＥＣＣ演算手段１０４２３は、これを用いて新パリティを得て、ディスクアレイコマンド関連付け管理手段１０４２４は、キュー中の、新パリティが得られた新パリティ用の子コマンドパケット＃４を複数個の関連付ける。そして、関連付けた新パリティ用の子コマンドパケットをディスクコマンド管理手段１０４３に送出する。

【０１７４】以降、前記第１実施例と同様に処理が行なわれ、すべての関連コマンドパケットに対応するデータ転送が終了したならば上記第１、２、３、４の実施例同様１回の割り込みがＩ／Ｏ処理装置から発行され、ディスクアレイコマンド関連付け管理手段１０４２４に処理が戻り、ディスクアレイコマンド関連付け管理手段１０４２４は、処理が終了した新パリティライト用のコマンドパケットの実行ステータスを正常終了とし、ディスクアドレス変換手段１０４２１に終了報告を行なう。

【０１７５】さて、このようにして、一つの親コマンドパケットに対応する４つの子コマンドパケットの終了報告が揃ったら、ディスクアドレス変換手段１０４２１は、前記第２実施例と同様に、その親コマンドパケット＃２の実行ステータスを正常終了とし、終了処理を行なう。ここで、ファイルアクセス管理部１０３、ライトアクセスの要求元のユーザプロセス１０２への終了報告は、既にすんでいるので行なわない。

【０１７６】以上のように、本第５実施例によれば、パリティの更新をデータの更新と分離し、パリティを遅延させて更新するような場合に、ＥＣＣ生成のキャッシュ利用による高速化により、ディスクアクセス性能を高速化することができると共に、ＣＰＵの割り込み処理を大幅に削減することができる。

【０１７７】なお、以上の各実施例では、Ｉ／Ｏ処理装置２において、関連コマンド終了判定手段２０３１はす

べての関連付けられたコマンドが終了するまでホスト１に終了報告を行なわないが、このようにすると、もし、関連付けられたコマンドの一つにエラーが生じた場合に、永遠に終了報告が延期されてしまう可能性がある。そこで、関連コマンド終了判定手段２０３１はあらかじめ設定した適当な時間でタイムアウトとし、関連付けられたコマンドコマンドの終了・未終了にかかわらずホストに終了報告を発行するようにしてもよい。

【０１７８】ホスト１のディスクコマンド管理部１０４はこの終了報告を受信すると、関連コマンドのステータスを確認し、未終了のコマンドがあったことを確認し、再度コマンドを発行するか、もしくは、コマンド実行失敗として要求元に報告する。

【０１７９】また、以上の各実施例は、ディスク装置ではなく、光ディスク装置、半導体記憶装置、磁気テープ装置、通信制御装置、等他のＩ／Ｏ処理を行う場合にも適用することができる。

【０１８０】

【発明の効果】以上のように、本発明によれば、ディスクアクセスに起因する、ＣＰＵへの割り込み発生回数を削減することにより、ＣＰＵの負荷を低減することのできる計算機システムを提供することができる。

【図面の簡単な説明】

【図１】本発明の第１実施例に係る計算機システムの構成を示すブロック図である。

【図２】本発明の実施例に係る計算機システムのディスクアクセスコマンド処理に関連する部位の構成を示したブロック図である。

【図３】本発明の第１実施例において用いるコマンドパケット、関連コマンド管理テーブルの構造および両者の関係を示した図である。

【図４】本発明の実施例に係るディスクコマンド管理手段の構成を示すブロック図である。

【図５】本発明の実施例に係る計算機システムのディスクアクセスの動作を示すタイムチャートである。

【図６】本発明の実施例に係るディスクアレイ管理手段の構成を示すブロック図である。

【図７】ディスクアレイの複数のディスクと、アクセス要求の関係を示す図である。

【図８】ＲＡＩＤにおけるデータとパリティの関係を示す図である。

【図９】本発明の第２実施例において用いるコマンドパケットと関連コマンド管理テーブルの構造および両者の関係を示した図である。

【図１０】本発明の第３実施例において用いるコマンドパケットと関連コマンド管理テーブルの構造および両者の関係を示した図である。

【図１１】本発明の第４実施例におけるディスクアクセスコマンドの経路を示す図である。

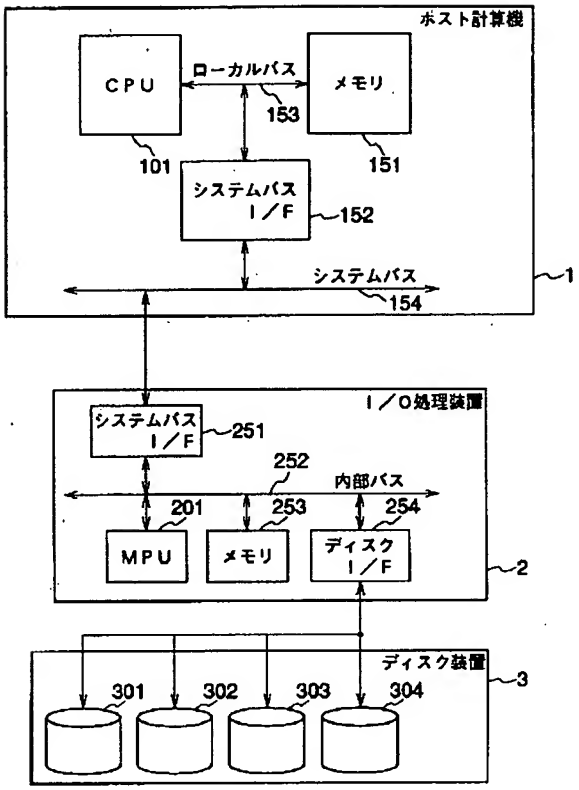
【図１２】本発明の第５実施例において用いるコマンドパケットと関連コマンド管理テーブルの構造および両者の関係を示した図である。

【符号の説明】

１・・・ホスト計算機
２・・・Ｉ／Ｏ処理装置
３・・・ディスクアレイ装置（又はディスク装置）
１０１・・・ＣＰＵ
１０２・・・ファイルアクセス要求生成部
１０３・・・ファイルアクセス要求管理部
１０４・・・ディスクコマンド管理部
１０５・・・Ｉ／Ｏ処理装置制御部
１５１・・・メモリ
２０１・・・ＭＰＵ
２０２・・・Ｉ／Ｏ処理装置Ｉ／Ｆ部
２０３・・・ディスクコマンド管理部
２０４・・・ディスクコマンド処理部
２０５・・・コマンド管理エリア
２５３・・・メモリ
１０００・・・ドライバ部
１０１０・・・オペレーティングシステム
１０４２・・・ディスクアレイ管理手段
１０４３・・・ディスクコマンド管理手段
１０４４・・・コマンド管理エリア
２０３０・・・関連コマンド登録手段
２０３１・・・関連コマンド終了判定手段
１０４３０・・・ディスクコマンド関連付け手段

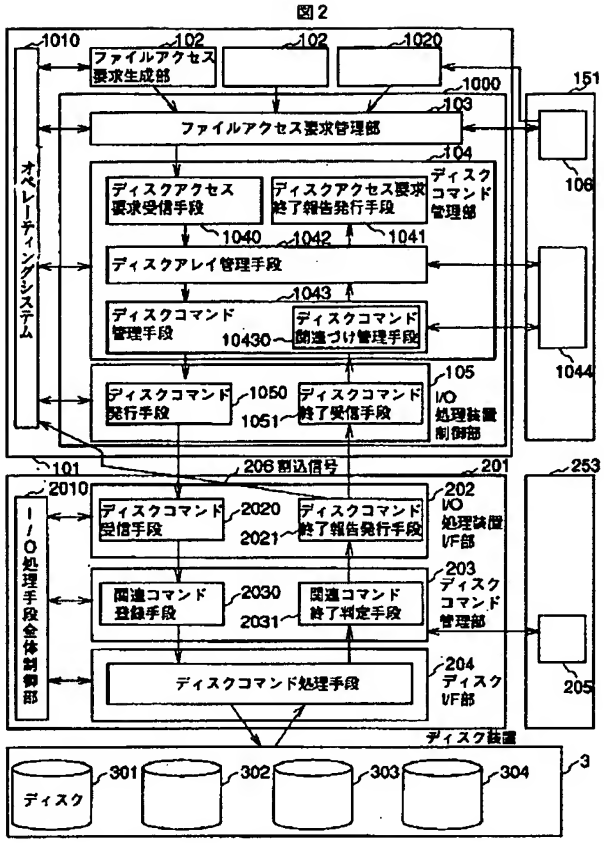
【図１】

図1



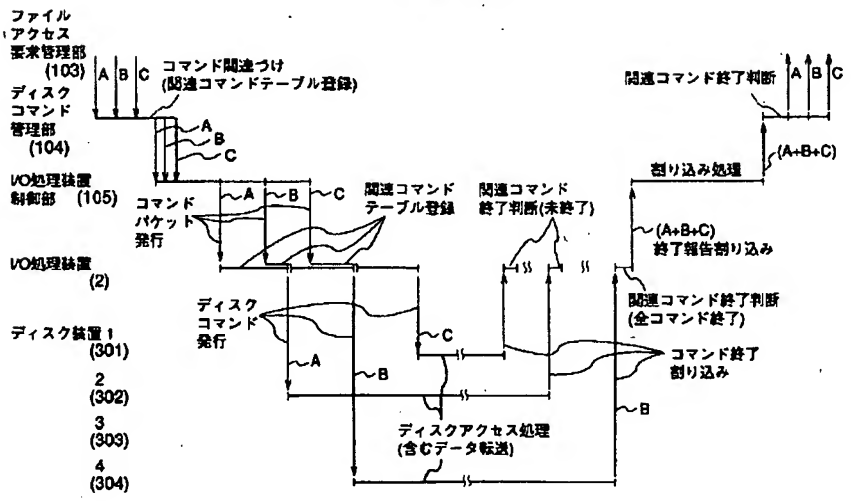
【図２】

図2



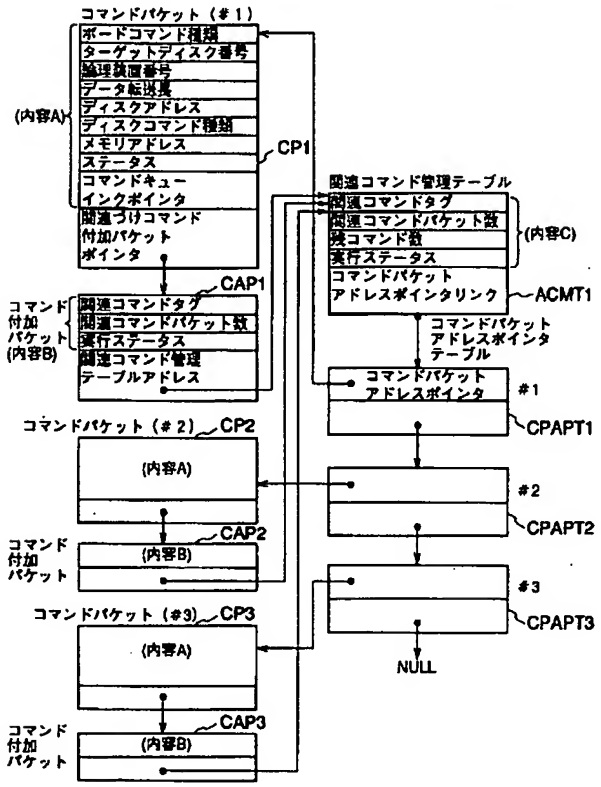
【図５】

図5
コマンド管理動作図



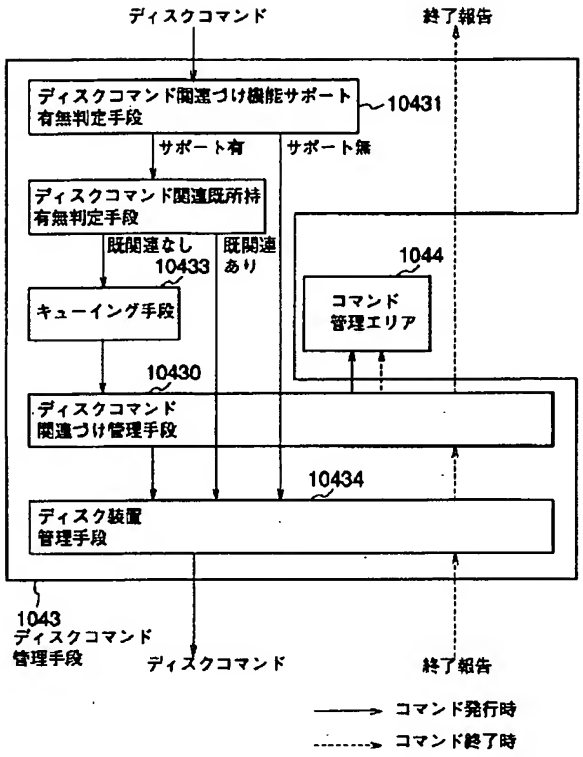
【図３】

図3
コマンドパケットの構造と関連コマンド管理テーブルの構造



【図４】

図4



【図７】

【図６】

図6

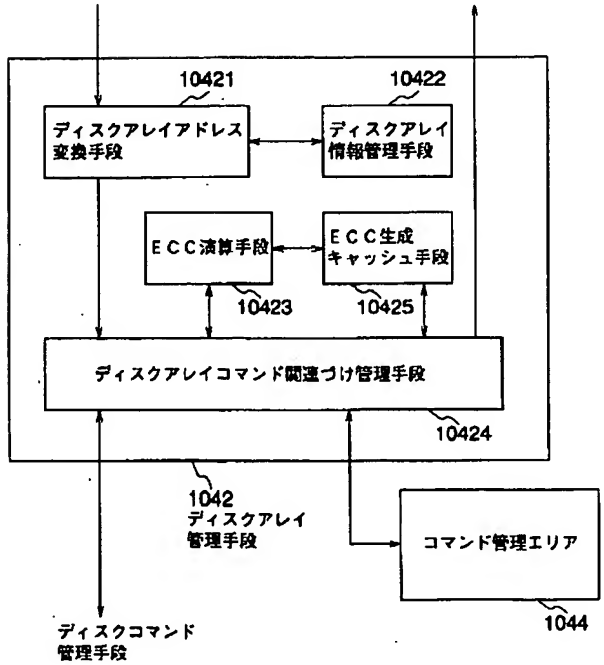
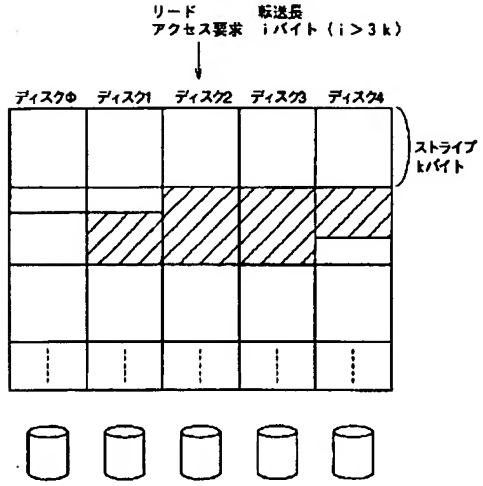
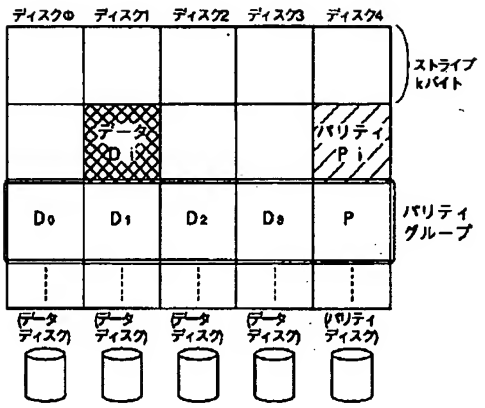


図7



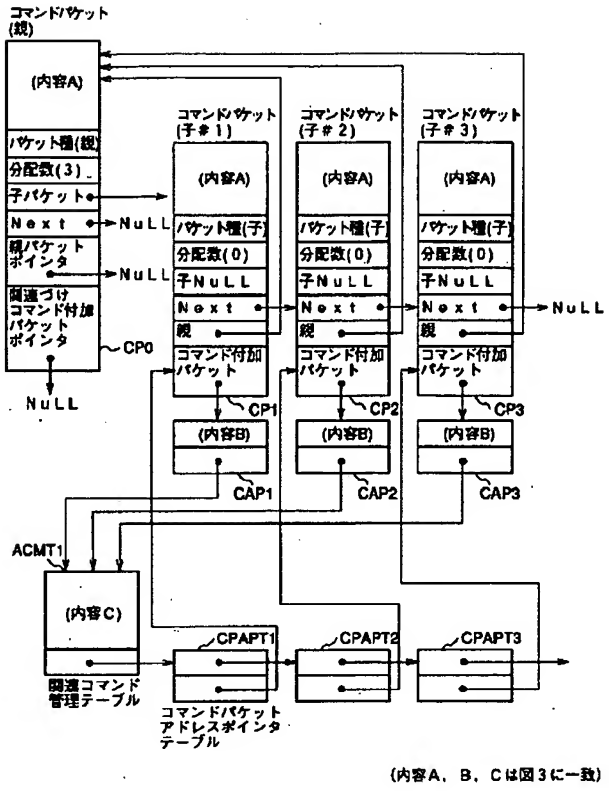
【図８】

図８


ディスク0
ディスク1
ディスク2
ディスク3
ディスク4
ストライプ kバイト
データ Di j
パリティ Pi
D0
D1
D2
D3
P
パリティグループ
(データディスク)
(データディスク)
(データディスク)
(データディスク)
(パリティディスク)

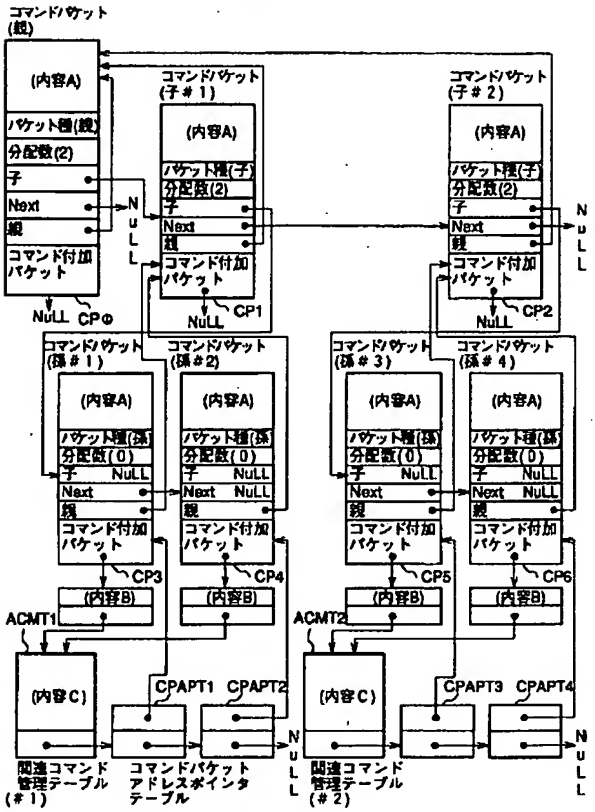
【図９】

図９


コマンドパケット(親)
(内容A)
パケット種(親)
分配数(3)
子パケット
Next
NULL
親パケットポインタ
NULL
関連づけコマンド付加パケットポインタ
CP0
NULL
コマンドパケット(子＃1)
コマンドパケット(子＃2)
コマンドパケット(子＃3)
パケット種(子)
分配数(0)
子NULL
Next
親
コマンド付加パケット
CP1
CP2
CP3
(内容B)
CAP1
CAP2
CAP3
ACMT1
(内容C)
関連コマンド管理テーブル
CPAPT1
CPAPT2
CPAPT3
コマンドパケットアドレスポインタテーブル
(内容A，B，Cは図３に一致)

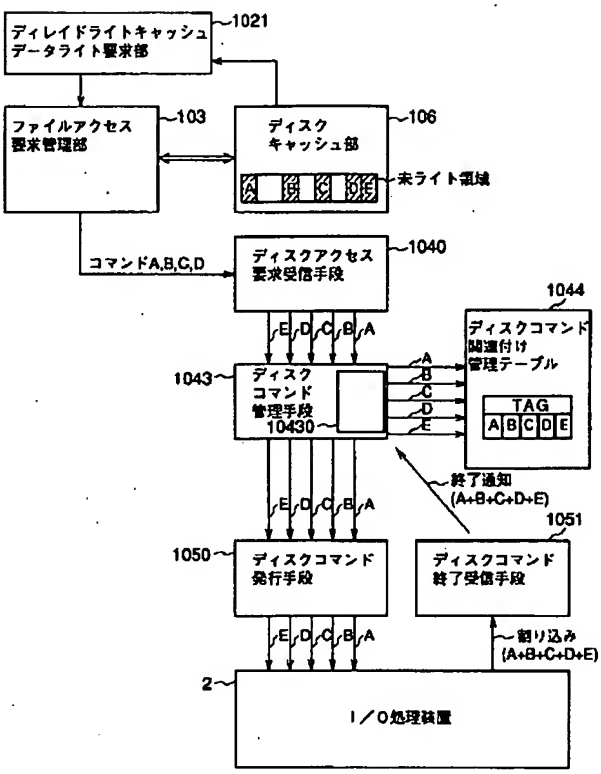
【図１０】

図１０

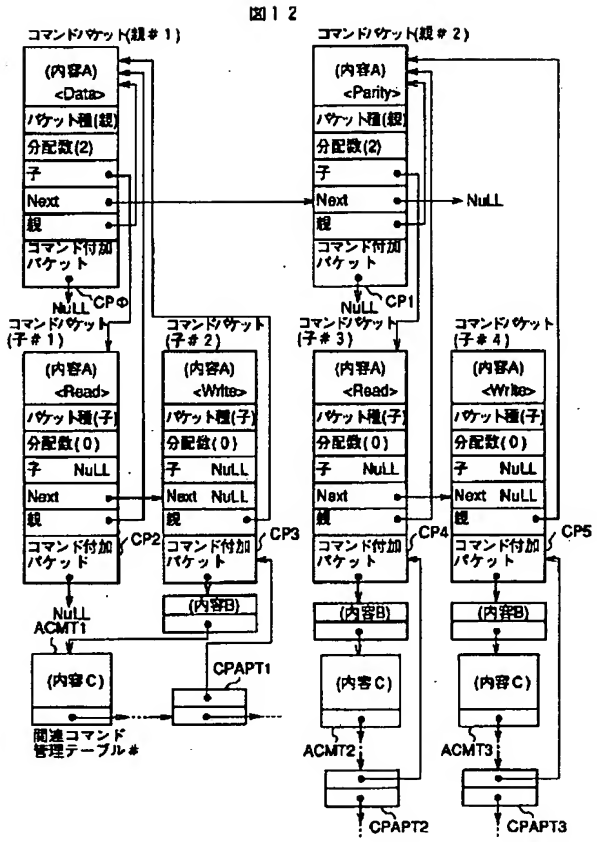


【図１１】

図１１
ライトキャッシュと関連づけ



【図１２】



フロントページの続き

(72)発明者　八木沢　育哉
神奈川県川崎市麻生区王禅寺1099　株式会社日立製作所システム開発研究所内

(72)発明者　大枝　高
神奈川県川崎市麻生区王禅寺1099　株式会社日立製作所システム開発研究所内

(72)発明者　荒川　敬史
神奈川県川崎市麻生区王禅寺1099　株式会社日立製作所システム開発研究所内

【公報種別】特許法第１７条の２の規定による補正の掲載
【部門区分】第６部門第３区分
【発行日】平成１１年（１９９９）７月３０日

【公開番号】特開平７―２７１５２１
【公開日】平成７年（１９９５）１０月２０日
【年通号数】公開特許公報７―２７１６
【出願番号】特願平６―５７１９７
【国際特許分類第６版】
G06F　3/06　540
　　13/10　340
【ＦＩ】
G06F　3/06　540
　　13/10　340 B

【手続補正書】
【提出日】平成１０年７月１日
【手続補正１】
【補正対象書類名】明細書
【補正対象項目名】請求項９
【補正方法】変更
【補正内容】
【請求項９】請求項８記載の計算機システムであって、前記ホスト装置のＣＰＵは、前記ディスクアレイより読み出したデータおよびパリティを、メモリ上に設けたキャッシュ領域に保持する手段を有し、
前記関連付け手段は、前記キャッシュ領域に、現データ読み出し用コマンドによって読み出すべきデータに対応するデータが存在する場合には、現データ読み出し用コマンドを前記入出力処理装置に渡さず、前記キャッシュ領域に、現パリティ読み出し用コマンドによって読み出すべきパリティに対応するパリティが存在する場合には、現パリティ読み出し用コマンドを前記入出力処理装置に渡さず、
前記パリティ生成部は、前記キャッシュ領域に、現データ読み出し用コマンドによって読み出すべきデータに対応するデータが存在する場合には、前記ディスクアレイより読み出した前記現データに代えて前記キャッシュ領域に存在するデータを用いて、前記現データと新データとの排他的論理和を生成し、前記キャッシュ領域に、現データ読み出し用コマンドによって読み出すべきパリティに対応するパリティが存在する場合には、前記ディスクアレイより読み出した前記現パリティに代えて前記キャッシュ領域に存在するパリティを用いて、現パリティと保持したパリティの排他的論理和を、前記新パリティ書き込み用コマンドによって書き込む新パリティとして生成することを特徴とする計算機システム。
【手続補正２】
【補正対象書類名】明細書
【補正対象項目名】請求項１０
【補正方法】変更
【補正内容】
【請求項１０】ホスト装置と補助記憶装置との間の入出力を担う入出力処理装置であって、
ホスト装置より渡された複数のコマンドに付加されている、当該コマンドの属するグループを示す情報に基づいて、渡された各コマンドと当該コマンドが属するグループの対応を表す情報を作成して記憶する手段と、ホスト装置より渡された各コマンドの指示する補助記憶装置のアクセスを実行する手段と、記憶した各コマンドと当該コマンドが属するグループの対応に基づいて、１つのグループに属する全てのコマンドの指示するアクセスが終了した場合に、当該グループについての処理が終了したことを前記ホスト装置に割り込みを用いて報告する手段とを有することを特徴とする入出力装置。